Journal of the Faculty of Commerce for Scientific Research©
Volum 56 – No. 3 – July 2019

Handling Missing Data in the Association-Marginal Model through Longitudinal Data Analysis: A Simulation Study[1]

Mahi Mohssen El-Zayat [2] Prof. Dr. Emtissal Mohamed[3]
Prof. Dr. Adel Halawa[4] Dr. Labiba El-Attar [5]

Faculty of Commerce - Alexandria University

ABSTRACT

Missing data can frequently occur in a longitudinal data analysis, where repeated measurements are taken over time. Unfortunately, missing data can lead to large standard errors in parameter estimates because nonresponse is compounded across times of data collection to produce small longitudinal sample sizes. Also, the problems of survey nonresponse (i.e., reduction in statistical power and threat of parameter bias) are a particularly salient challenge for longitudinal researchers. Thus, the main goal of this paper is to introduce a new idea that describes simultaneously the association structure (A) with the marginal distributions (M) of the responses for longitudinal data in the presence of missing data (MS), through a composite link. This new idea (AM-MS) is of great

[1] Received in 2/6/2019, accepeted in 13/7/2019. This artiocle is derived from Ph.D. desertation by Mahi Mohssen and supervised by Prof. Emtissal Mohamed, Prof. Adel Halawa and Dr. Labiba El-Attar.

[2] Mahi Mohssen Younes Mohamed El-Zayat Assistant Lecturer at Statistics, Mathematics and Insurance Department, Faculty of Commerce-Alexandria University.
(E-mail: stat4statalxx@gmail.com)

[3] Professor of Statitics at Statitics, Mathemeatics and Insurance Department, Faculty of Commerce, Alexandria University.

[4] Professor of Statitics at Statitics, Mathemeatics and Insurance Department, Faculty of Commerce, Alexandria University.

[5] Associate Professor of Statitics at Statitics, Mathemeatics and Insurance Department, Faculty of Commerce, Alexandria University

importance where it is applicable for large and sparse tables. In addition, it can also be used for fitting log linear models to contingency tables with missing data (MS) and fitting models with various assumptions about the missing data mechanisms either MCAR, MAR or NMAR. A simulation study will be performed to apply this new idea, under various situations including (missing mechanisms, missing rates and five methods for handling missing data). The goodness-of-fit test statistics and the number of adjusted residuals greater than 2 are used as evaluation criteria.

Keywords: Association model (A) - Marginal model (M) - Simultaneous AM model - Missing data (MS) -Ordinal data -Composite link function -Generalized linear models (GLM) – CC - mode imputation – LOCF – KNNI – MI - Longitudinal studies.

1-INTRODUCTION

Nakai et al. (2014) and Nooraee et al. (2018) indicated that longitudinal studies are an important source of information in health sciences and other areas but often have the problem of missing data. Missing values in longitudinal studies occur when not all of the planned measurements of a subject outcome vector are actually observed. Thus, in analyzing data from clinical trials and longitudinal studies, missing data may appear. Missing data could introduce bias and lead to erroneous statistical inferences. Thus, whenever there are missing data, there is loss of information, which causes a reduction in efficiency, decrease statistical power and a drop in the precision in statistical inference. So, missing data should not be ignored and should be handled.

Lang & Agresti (1994) indicated that the analysis of complete longitudinal multivariate categorical response data is very common and useful in a variety of applications, especially for social studies. The longitudinal multivariate categorical responses are obtained from repeated measurements taken on subjects over time or occasions. These responses are often inevitably interrelated and the purpose of their modeling is to describe the association structure (changes at the individual level from one point to another) among these responses and also to know the behavior of their marginal distributions (changes within the year for different individuals). The Generalized Linear Models (GLM) are usually used for this analysis. Most of these models allow the researchers to model the association structure among these responses or to model their marginal distributions separately.

Lang et al. (1997) modeled simultaneously the association structure (A) with the marginal distributions (M) of the responses using a composite link function. This composite link lies between the two models of the log linear model that describes the association structure between the variables and the logistic regression model that describes the marginal distributions of the responses.The model derived from these two models is known as the simultaneous Association Marginal (AM) model which contains a composite link function that consists of the log and the logit links. This AM model provides improved model parsimony, one also obtains a single test that summarizes goodness of fit and a single set of fitted values and residuals. Also, estimators of the simultaneous AM model parameters are more efficient than with separate fitting process procedures.

Lang and Agresti (1994); Lang et al, (1997) and Lang and Eliason, (1997) are the researchers whom introduced the AM model, conditioned that all the data should be observed without any missing values, but missing data (MS) are often a problem for multivariate longitudinal response data.

Missing data is vital subject to perform a proper longitudinal analysis. Some researchers ignore and discard all missing data to have complete dataset. However, it can result in a very substantial loss of information. Thus, this paper will introduce the new idea (AM-MS) that can be used to simultaneously describe the association structure (A) with the marginal distributions (M) of the responses in the presence of missing data (MS).

2- MECHANISMS AND METHODS FOR HANDLING MISSING DATA

2-1 Missing Data Mechanisms

Rubin, (1976) defined a clear classification of missingness that has become the standard for any discussion of this topic. Rubin classified missing data mechanisms into three different types:

2-1-1 Missing completely at random (MCAR)

MCAR data are missing due to a completely random process. Suppose variable Y has some missing values. We will say that these values are MCAR if the probability of missing data on Y is unrelated to the value of Y itself or to the values of any other variable in the data set (Allison, 2002 & Bori, 2013).

The MCAR assumption is defined as:

$$P\,(Y \text{ missing} \mid Y, X) = P(Y \text{ missing}).$$

This assumption states that missingness is not related to any factor, known or unknown in the study, (i.e. missingness is unrelated to the data).

2-1-2 Missing at random (MAR)

It is a weaker assumption than MCAR. This assumption states that:

$$P(Y\ missing\ |Y, X) = P(Y\ missing\ |X)$$

Horton and Kleinman, (2007) described MAR mechanism and stated that the missingness depends only on observed quantities, which may include outcomes and predictors. Thus, the missing value of the variable of interest does not depend on the variable of interest but conditionally depends on other observed variables' values.

2-1-3 Not missing at random (NMAR)

This is the case in which the probability of missingness for the variable of interest depends upon the value of that variable itself. For example; there is a high rate of missing data on an item asking about participants' annual income. It may be the case that participants with high rates of income are more likely to omit this item because they are uncomfortable with others knowing their income. The student may have a missing value from any grade of his grades when filling a survey because he does not want to tell his bad grade to anyone.

2-2 Methods for Handling Missing Data

Bori, (2013) & Rithy, (2016) indicated that the development of imputation methods for handling and analyzing data with missing values has been an active area of research. Hence, missing data must be examined carefully and the proper imputation method should be used. This paper focused on the comparison among five selected imputation methods according to the nature of longitudinal ordinal data, these methods are: Complete Case Analysis (CC), mode imputation, last observation carried forward (LOCF), K-nearest neighborhood imputation (KNNI) and finally multiple imputation (MI).

2-2-1 Complete Case Analysis (CC)

This method deletes all cases with missing data and then performs statistical analyses on the remaining complete data set (which has a smaller sample size). Since all cases containing missing data have been removed, there is no missing data problem to handle. Therefore, all statistical methods can be used to analyze the smaller data set.

Zhu, (2014); Nakai et al., (2014) & Al-Zahrani, (2018) indicated that one major advantage of this method is its ease of use. In fact, virtually all statistical programs inc-orporate this method as a default method because it accommodates any type of statistical analysis. The method may be preferred under the situation in which the sample size is large, the proportion of missing data is small, and the missing data mechanism is MCAR. For MCAR missing data, the method will yield unbiased parameter estimates. (Nakai et al., 2014) indicated that this method works well when the data is MCAR, which rarely happens in reality.

While the disadvantages of this approach are that it results in loss of information because a large part of the original sample is excluded and it could possibly lead to losing statistical power due to the reduction of the sample size. Also, complete case techniques decrease the efficiency such that the variation (i.e., the standard error) around the true estimate is too large.

2-2-2 Mode Imputation

Baraldi & Enders, (2010) indicated that mode imputation method replaces missing values of a categorical variable by the mode of non-missing cases of that variable. Mode imputation is used when the missing mechanism is MCAR. It is one of the easiest ways in the case of categorical data is to fill in each missing value with the mode of observed values.This is a common practice; nonetheless, the major disadvantage of mode imputation is that it creates spikes in the distribution by concentrating all the imputed values in the mode.This is a single imputation method, since only one value is used to replace each missing observation.

2-2-3 Last Observation Carried Forward (LOCF)

Al-Zahrani, (2018) indicated that LOCF method is considered as the simplest imputation approach and can only be applied under a longitudinal study with MCAR mechanism. In this method the missing values are replaced by the last observed value from that variable. The advantage of this method is easy to understand and popular for handling missing data. Also, unlike the listwise deletion method, the sample size does not change. While the disadvantage of this method is that, it can bias results and lead to either overestimation or underestimation of the parameter estimates.

2-2-4 K-Nearest Neighborhood Imputation (KNNI)

Schlomer et al., (2010) indicated that KNN imputation method uses the Knearest neighbors approach to impute missing values. What KNN imputation does in simpler terms is as follows: For every observation to be imputed, it identifies 'K' closest observations based on the euclidean distance and computes the weighted average (weighted based on distance) of these 'K' observations. The advantage is that you could impute all the missing values in all variables with one call to the function. It takes the whole data frame as the argument and you don't even have to specify which variable you want to impute.

2-2-5 Multiple Imputation (MI)

Rubin, (1987) was the first to propose multiple imputation to analyze incomplete data under the MAR mechanism. (Rithy, (2016) & Zhu, 2014) indicated that multiple imputation (MI) creates several iterations of imputed datasets and condenses all datasets into one. The idea of MI is to impute a missing value multiple times and hence generates multiple (m) data sets. Then, these imputed data sets are analyzed separately by standard procedures that are commonly used in analyzing complete data sets. Finally, the results of analyses are combined to get a final set of parameter estimates.

Allison, (2002); Garg, (2013) and Kombo et al., (2017) indicated to the major advantage of MI is that it allows the use of complete-data methods for data analysis and incorporating random errors in the imputation process. In addition, MI increases the efficiency of the estimates through minimizing the standard errors, which are used for significance testing and/or construction of confidence intervals around these parameter estimates. Finally, the MI procedure provides accurate standard errors and therefore accurate inferential conclusions. So, the precision of parameter estimates and accuracy of standard errors make MI one of the best options for handling missing data under MAR mechanism.

Deng et al., (2016) pointed out that MI needs more effort to create the multiple imputations, more time to run the analyses, and more computer storage space for the imputation-created data sets.

3- Modeling the AM Model with MS Model (AM-MS)

In this subsection a new model (AM-MS) will be introduced that simultaneously describes the association structure (A) with the marginal distributions (M) of the responses when the data contain missing values (MS). This new model will combine the A model with the M model in the presence of MS through a composite link. The new model consists of the following three models:

First: The association (A) model which is specified as a standard log linear model for the vector of expected counts m:

A: $\log m = X_1\beta_1$ (1)

where X_1 is a matrix of known constants and is assumed to be of full column rank P_1 and β_1 is a P_1 x1 vector of parameters.

Second: The marginal (M) model can be written as generalized log linear model,

$$\text{M: } C_2 \log M_2 m = X_2 \beta_2 \quad (2)$$

where C_2, M_2 and X_2 are matrices of known constants. C_2 is of full row rank and since the analysis depends on the log linear models for expected counts, then C_2 is the identity matrix.The elements of M_2 are non negative and consist of ones and zeros, such that $M_2 m$ is a vector that can be partitioned as linear combinations of components of m_1,linear combinations of components of m_2, ... and linear combinations of components of m_K .The linear combinations of the expected cell counts of m are within each covariate pattern (inside the same pattern) not across different patterns. The matrix X_2 is of full column rank P_2 and β_2 is a P_2 x 1 vector of parameters.

Third: The missing data (MS) model which is specified as a saturated log linear for the vector of expected counts m:

$$\text{MS: } \log M_3 m = X_3 \beta_3 \quad (3)$$

where M_3 is a matrix of 0's and 1's, that tells which elements of the unobserved vector m are summed to result in an estimated observed frequency, X_3 is a matrix of known constants and is assumed to be of full column rank P_3 and β_3 is a P_3 x1 vector of parameters.

Thus, by combining the previous three models, the new multinomial and Poisson AM-MS are respectively:

$$\text{AM-MS: } C \log Mm = X\beta \text{, samp (m) = 0} \quad (4)$$

and

$$\text{AM-MS: } C \log Mm = X\beta \quad (5)$$

For the previous two models, $C = \oplus_{i=1}^{3} C_i$ is a block diagonal matrix with C_1, C_2 and C_3 as the blocks, $M' = (M_1', M_2', M_3')$, $m = (m_1, m_2, ...,m_K)'$, $X = \oplus_{i=1}^{3} X_i$ is a block diagonal matrix with X_1, X_2 and X_3 as the blocks, the symbol $\oplus$ is the direct sum operator, $\beta = (\beta_1', \beta_2', \beta_3')'$ and $samp(m) = 0$ denotes the multinomial identifiability constraints. For the case of K independent multinomial samples of sizes $n = (n_1, n_2, ..., n_K)'$, these constraints have the form $Z'm - n = 0$, where the matrix

$Z = \oplus_{k=1}^{K} 1_r$, 1_r is an r - dimensional vector of one's and r denotes the number of response profiles for each covariate pattern of K identical components. These multinomial sampling constraint simply state that within one of the K covariate patterns, say level k, the sum of expected counts (m_k) should be equal to n_k.

Thus, by combining the AM model with the MS model the simultaneous AM models will be applicable for large and sparse tables with MS. Also, these simultaneous AM model with MS can be used in fitting log linear models to contingency tables with missing data (MS). Besides the previous advantages, this new model (AM-MS) can be used for fitting AM models with MS by assuming various assumptions about the missing data mechanisms (MCAR, MAR or NMAR) and different missing rates. Also, AM-MS can be used for comparing AM models after applying the different methods for handling MS to choose the best method for treating MS in the AM model in each missing mechanism with each missing rate.

4- DESIGN OF THE SIMULATION STUDY

To achieve the research's goal, a simulation study was performed to simulate four time points (responses) each with three levels (J = 3) using the SimCorMultRes package version 1.4.1 in R. Touloumis, (2016). & (2018). indicated that this package is the first R package that targets specifically on the generation of correlated binary, nominal or ordinal responses under marginal model specification.

The rmul.clm function in the SimCorMultRes package was used to generate a longitudinal ordinal data Y_{it} (i=1,2,....,N, t=1,... ,T) for i-th subject at t-th occasion. The simulation of the data was conducted according to a cumulative logit model:

$$\text{logit}\,[P(Y \leq j)] = \alpha_j + \beta x, \qquad (6)$$

where α_j is the intercept for level j and β is the slope when using one explanatory variable, x. Here in this paper each response has J = 3 categories, then there will be 2 intercepts only (α_j = 0.5, 1.5) since models for cumulative probabilities do not use the final one, P (Y ≤ J), since it necessarily equals 1. In this model the parameter β which is the slope (β = 1.5) describes the effect of X on the log odds of response in category j or below. In the model formula, β does not have a j subscript; this means that the model assumes an identical effect of X for all J-1 logits. The intercepts and the slope are assumed to be constant through the simulation study. The correlation coefficient (ρ) between the responses is a positive correlation coefficient with values of 0.2 or 0.6, to see the effect of low and high correlation between the responses. The total number of cases or subjects (N) simulated is 200 and there are 4 measurements (points in time) for each subject. Thus, there are here two scenarios for the complete data, either:

1) Scenario 1 (N = 200, J = 3 and $\rho = 0.2$) or
2) Scenario 2 (N = 200, J = 3 and $\rho = 0.6$).

Then, missing data will be inserted and injected in each scenario using the three missing mechanisms (MCAR, MAR and NMAR). In addition, without loss of generality, the missing pattern was assumed to be arbitrary, where missingness can occur at any point in time and to any subject.The missing rate was assumed to be either low missing rate (10%) or high missing rate (50%) with MCAR and NMAR. Finally, the performance of five methods (that is, CC analysis, mode imputation, LOCF, KNNI and MI) for handling MS in the AM model was compared based on three goodness-of-fit test statistics, and the number of adjusted residuals greater than 2 for each AM model. The three goodness-of-fit test statistics are: G2 which is the likelihood ratio statistic, $\chi 2$ which is Pearson's Score statistic, W2 which is the generalized Wald statistic.

The goal of this paper is to select the appropriate missing data imputation method for handling MS then to estimate the AM model after handling the missing values, to choose the best method for handling MS and its effect on the AM model. Thus, the best method for handling MS in the AM model is the method which leads to an AM model with the smallest values of the test statistics and the smallest number of adjusted residuals greater than 2.

Thus, the AM model will be estimated using the following cases; for each of MCAR and NMAR there will be 10 different cases: 2 (missing rates) ×5 (methods for handling MS). While for MAR there will be only the 5 methods for handling MS. Here there will be 25 cases for each scenario; 10 for MCAR, 10 for NMAR and 5 for MAR. Therefore, there will be 50 different cases for both scenarios.

5- SIMULATION RESULTS

This section reports the results of the simulation study comparing the effect of the methods CC, mode imputation, LOCF, KNNI and MI on handling MS in the AM model.

5-1 MCAR Simulation Results

Table 1 and Table 2 show the simulation results of scenario 1 and scenario 2 respectively, for estimating the AM model with MCAR missing mechanism and low (10%) missing rate using: CC, mode imputation, LOCF, KNNI and MI.

Thus, after handling MS in the AM model and depending on the goodness-of-fit test statistics and the number of adjusted residuals greater than 2 as evaluation criteria, the best method for handling MS in scenario 1 (Table 1) is LOCF while the worst method for handling MS is KNNI. Also, the best method for handling MS in scenario 2 (Table 2) is also the LOCF while the worst method for handling MS in scenario 2 is KNNI method.

Therefore, either with low or high correlation structure, the best method for handling MS in the AM model is LOCF, while the worst method is KNNI.

Table 3 and Table 4 show the simulation results of scenario 1 and scenario 2 respectively, for estimating the AM model with MCAR missing mechanism and high (50%) missing rate using: CC, mode imputation, LOCF, KNNI and MI.

Thus, after handling MS in the AM model and depending on the goodness-of-fit test statistics and the number of adjusted residuals greater than 2 as evaluation criteria, the best method for handling MS in scenario 1 (Table 3) is mode imputation method, while the worst method for handling MS is MI. While for scenario 2 (Table 4), none of the methods lead to a significant AM model, where all the method leads to a non significant AM model.

Table 1: Estimates of the AM model for Scenario 1with MCAR and 10% missing rate, using: CC, mode imputation, LOCF, KNNI and MI:

	CC	Mode imputation	LOCF	KNNI	MI
Intercept	-3.8116	-3.1954	-3.2252	-3.7078	-3.4958
Y12	-3.3539	-3.5796	-3.2056	-3.9217	-3.1959
Y13	-6.0538	-5.7598	-5.0106	-6.9313	-5.2559
Y22	-2.7925	-2.6052	-2.4696	-2.8127	-2.8275
Y23	-4.5744	-3.3274	-3.2665	-4.2060	-4.2984

Y32	-3.4551	-3.0619	2.7779	-3.1765	-2.8286
Y33	-6.1555	-4.4419	-3.9423	-2.8127	-4.3424
Y42	-2.5358	-3.0109	-2.8160	-3.2654	-2.8775
Y43	-4.0783	-4.3323	-4.0082	-5.3427	-4.5264
Y1score:Y2score	0.4604	0.2546	0.3908	0.4515	0.2981
Y1score:Y3score	0.6610	0.6391	0.4619	0.6328	0.4336
Y1score:Y4score	0.3761	0.5235	0.3758	0.5966	0.5152
Y2score:Y3score	0.3260	0.1455	0.1506	0.1481	0.3228
Y2score:Y4score	0.2364	0.2846	0.2171	0.3736	0.3321
Y3score:Y4score	0.3872	0.1566	0.2839	0.3176	0.2318
CUT1	0.1943	0.4500	0.2652	0.3846	0.2741
CUT2	1.0108	1.0953	0.9407	1.0698	1.0079
RESPY2	0.1046	0.0744	0.0340	-0.0774	0.0741
RESPY3	0.1985	0.1513	0.1404	0.1031	0.0289
RESPY4	-0.1046	0.0606	0.1364	-0.0680	-0.0531
G2	81.429 (0.1454)	71.64 (0.3903)	70.84 (0.42)	85.99 (0.081)	82.141 (0.1334)
χ2	86.415 (0.0765)	80.542 (0.162)	66.25 (0.57)	88.099 (0.060)	80.045 (0.171)
W2	34.399 (0.9998)	38.84632 (0.998)	30.193 (1)	39.938 (0.998)	35.587 (0.9997)
adj.resd. > 2	7	4	2	8	4

Remark: Y12....Y44: the main effects association terms, Y1score:Y2score is a linear-by-linear association term, α_1 = CUT1, α_2 =CUT2, RESPY2 is the value of the second response in the marginal model, G2 is the likelihood ratio statistic, χ2 is Pearson's Score statistic, W2is generalized Wald statistic, (*) is the p-value corresponding each test statistic and adj. resd are the number of adjusted residuals which are greater than 2 in each case.

Table 2: Estimates of the AM model for Scenario 2 with MCAR and 10% missing rate, using: CC, mode imputation, LOCF, KNNI and MI:

	CC	Mode imputation	LOCF	KNNI	MI
Intercept	-5.2198	-4.1503	-4.0475	-4.9731	-4.8421
Y12	-5.2980	-4.3215	-3.9563	-4.5179	-4.7260
Y13	-11.3405	-8.0724	-7.2124	-8.8965	-9.6852
Y22	-4.2987	-3.6300	-3.4745	-4.4823	-3.7519
Y23	-8.5971	-6.0338	-5.8591	-8.8188	-6.9850
Y32	-4.1357	-3.7941	-3.4501	-4.1794	-3.8684
Y33	-8.2136	-6.6572	-5.8408	-7.9724	-7.3554
Y42	-3.6407	-3.6852	-3.4984	-4.1123	-3.9723
Y43	-6.8000	-6.2629	-5.8658	-7.7819	-7.6474
Y1score:Y2score	0.8366	0.2899	0.5990	0.6101	0.4867
Y1score:Y3score	1.0824	0.9688	0.6849	0.9361	0.9007
Y1score:Y4score	0.8327	0.7738	0.5394	0.6096	0.9791
Y2score:Y3score	0.6702	0.5067	0.3334	0.6354	0.6016
Y2score:Y4score	0.5827	0.5761	0.4353	0.9147	0.5899
Y3score:Y4score	0.2419	0.0905	0.3622	0.3517	0.2737
CUT1	0.3693	0.4489	0.2687	0.4060	0.3398
CUT2	1.0704	1.0963	0.9347	1.0433	1.0479
RESPY2	-0.0337	0.1438	0.1022	-0.0201	-0.0179
RESPY3	-0.0374	0.0598	0.0675	-0.0093	-0.0207
RESPY4	-0.0771	0.1109	0.1526	-0.0146	0.0101
G2	57.4149 (.8389)	82.1803 (0.1327)	58.1003 (0.823)	77.98325 (0.2148)	74.8433 (0.2945)
$\chi 2$	65.4622 (0.5985)	78.6150 (0.2006)	47.292 (0.9788)	86.82409 (0.0723)	76.79798 (0.243)
W2	25.97029 (1)	36.3939 (0.9996)	22.9725 (1)	35.74292 (0.9997)	34.81793 (0.9998)
adj. resd. > 2	6	8	0	5	4

Remark: Y12....Y44: the main effects association terms, Y1score:Y2score is a linear-by-linear association term, α_1= CUT1, α_2=CUT2, RESPY2 is the value of the second response in the marginal model, G2 is the likelihood ratio statistic, $\chi 2$ is Pearson's Score statistic, W2is generalized Wald statistic, (*) is the p-value corresponding each test statistic and adj. resd are the number of adjusted residuals which are greater than 2 in each case.

Table 3: Estimates of the AM model for Scenario1with MCAR and 50% missing rate, using: CC, mode imputation, LOCF, KNNI and MI:

	CC	Mode imputation	LOCF	KNNI	MI
Intercept	-5.9248	-2.1468	-2.7016	-4.2556	-3.7479
Y12	-3.5319	-3.4721	-2.2802	-3.9841	-3.6262
Y13	-7.5713	-4.3300	-2.5709	-6.6367	-6.5537
Y22	-4.156	-3.0265	-2.0144	-4.4007	-2.7432
Y23	-8.5411	-3.4060	-1.9102	-7.6301	-4.4056
Y32	-5.4224	-3.0311	-2.2546	-4.5518	-2.6392
Y33	-12.0285	-3.3787	-2. 5172	-8.2448	-4.1763
Y42	-3.4451	-3.1260	-2.4417	-4.0814	-2.7602
Y43	-7.7785	-3.7135	-2.8204	-6.8760	-4.5130
Y1score:Y2score	0.0534	0.1527	0.0991	0.3803	0.6166
Y1score:Y3score	0.9373	0.3224	0.2834	0.6668	0.6636
Y1score:Y4score	0. 7179	0.3617	0.1793	0.5317	0.3622
Y2score:Y3score	0.6609	0.1489	0.0472	0.8170	-0.1290
Y2score:Y4score	0.9296	0.2776	0.1679	0.5018	0.4392
Y3score:Y4score	1.0739	0.0813	0.2082	0.5976	0.3450
CUT1	-0.0643	1.3417	0.2243	0.6778	0.1342
CUT2	1.0347	1.9185	0.8411	1.2225	0.9196
RESPY2	0.4195	-0.0978	0.1821	0.1559	0.1610
RESPY3	0.1471	-0.0425	0.0151	-0.0267	0.1373
RESPY4	-0.8003	-0.1865	0.2055	0.0237	-0.1231
G2	37.872 (0.99)	53.27 (0.9191)	105.878 (0.0029)	126.18 (3.244e-05)	197.99 (2.154e-14)
χ^2	95.064 (0.02)	58.79 (0.80)	127.61 (2.283e-05)	166.61 (4.777e-10)	236.6779 (0)
W2	15.018 (1)	26.74 (1)	49.939 (0.9594)	48.16621 (0.9734)	69.47938 (0.4612)
adj. resd. > 2	5	4	8	10	14

Remark: Y12….Y44: the main effects association terms, Y1score:Y2score is a linearby-linear association term, α_1= CUT1, α_2=CUT2, RESPY2 is the value of the second response in the marginal model, G2 is the likelihood ratio statistic, $\chi 2$ is Pearson's Score statistic, W2is generalized Wald statistic, (*) is the p-value corresponding each test statistic and adj. resd are the number of adjusted residuals which are greater than 2 in each case.

Table 4: Estimates of the AM model for Scenario2 with MCAR and 50% missing rate, using: CC, mode imputation, LOCF, KNNI and MI:

	CC	Mode imputation	LOCF	KNNI	MI
Intercept	-------------	-2.8256	-2.9875	-6.8209	-4.9630
Y12	-------------	-3.3395	-2.6346	-5.5659	-3.5130
Y13	-------------	-4.2433	-3.7523	-11.9706	-6.1181
Y22	--------------	-4.1562	-2.1845	-7.2112	-4.8244
Y23	---------------	-6.0404	-2.5999	-16.8096	-9.5225
Y32	----------------	-3.6799	-2.2489	-5.2446	-4.9344
Y33	----------------	-5.0852	-2.8777	-10.8298	-10.0992
Y42	-------------	-3.6382	-2.8777	-4.9687	-3.8785
Y43	-------------	-4.7628	-3.7496	-10.2600	-7.0287
Y1score:Y2score	---------------	0.0248	0.1476	1.5532	-0.2562
Y1score:Y3score	--------------	0.4151	0.3428	0.8915	0.8350
Y1score:Y4score	---------------	0.4572	0.4116	0.4996	0.8233
Y2score:Y3score	--------------	0.7435	0.1541	1.0071	1.6393
Y2score:Y4score	---------------	0.5295	0.1851	1.5037	0.6328
Y3score:Y4score	--------------	-0.0398	0.1568	0.6529	0.2828
CUT1	-------------	1.1084	0.1868	0.5402	0.3059
CUT2	--------------	1.6633	0.8909	1.0791	0.9642
RESPY2	---------------	0.2771	0.2232	0.0308	0.2451
RESPY3	--------------	0.0683	-0.0239	0.0477	-0.0942
RESPY4	---------------	0.2410	0.3123	-0.0903	-0.0043
G2 χ^2 W2	------------------ ----------------- ------------------	86.11542 (0.0797) 155.114 (1.431e-08) 53.9829 (0.9078)	103.978 (0.00415) 104.962 (0.00343) 42.41676 (0.9951)	93.7281 (0.0256) 117.297 (0.000258) 34.23459 (0.9999)	175.076 (3.535e-11) 293. 91 (0) 59.05125 (0.7978)
adj. resd. > 2	---------	13	9	11	10

Remark: Y12....Y44: the main effects association terms, Y1score:Y2score is a linear-by-linear association term, α_1 = CUT1, α_2 =CUT2, RESPY2 is the value of the second response in the marginal model, G2 is the likelihood ratio statistic, $\chi 2$ is Pearson's Score statistic, W2is generalized Wald statistic, (*) is the p-value corresponding each test statistic and adj. resd are the number of adjusted residuals which are greater than 2 in each case.

5-2 MAR Simulation Results

Table 5 and Table 6 show the simulation results of scenario 1 and scenario 2 respectively, for estimating the AM model with MAR missing mechanism using: CC, mode imputation, LOCF, KNNI and MI.

Thus, after handling MS in the AM model and depending on the goodness-of-fit test statistics and the number of adjusted residuals greater than 2 as evaluation criteria, the best method for handling MS in scenario 1 (Table 5) is MI while the worst method for handling MS is KNNI. Also, the best method for handling MS in scenario 2 (Table 6) is also the MI while the worst method for handling MS in scenario 2 is KNNI method.

Therefore, either with low (scenario 1) or high (scenario 2) correlation structure, the best method for handling MS in the AM model is MI, while the worst method is KNNI.

Table 5: Estimates of the AM model for Scenario1 with MAR mechanism using: CC, mode imputation, LOCF, KNNI and MI:

	CC	Mode imputation	LOCF	KNNI	MI
Intercept	-3.5362	-3.3561	-3.4062	-4.0393	-3.5567
Y12	-3.3260	-3.3672	-2.9068	-3.6134	-3.5670
Y13	-5.5342	-5.4477	-4.5784	-6.5541	-6.1177
Y22	-2.6110	-3.1759	-2.3082	-3.1662	-2.4544
Y23	-3.7941	-4.6885	-3.1876	-5.5435	-3.4550
Y32	-3.0907	-2.7482	-3.0337	-3.1254	-2.9114
Y33	-4.8206	-4.0122	-4.8178	-5.2931	-4.5178
Y42	-2.9615	-3.5415	-2.5994	-3.2146	-2.9705
Y43	-4.6255	-5.5393	-3.7843	-5.7177	-4.6942
Y1score:Y2score	0.1923	0.1090	-0.0445	0.2461	0.2516
Y1score:Y3score	0.5981	0.7551	0.7549	0.5456	0.6420
Y1score:Y4score	0.5661	0.5206	0.3410	0.7383	0.5671
Y2score:Y3score	0.3213	0.1890	0.3015	0.5530	0.1838
Y2score:Y4score	0.3885	0.6969	0.4916	0.5852	0.3606
Y3score:Y4score	0.1152	-0.0563	0.0226	0.0811	0.1790
CUT1	0.2501	0.3316	0.3096	0.2932	0.3144
CUT2	0.9595	1.0065	1.0395	1.0680	1.0333
RESPY2	-0.0657	0.4478	-0.2126	-0.1849	-0.1252
RESPY3	0.1990	0.0407	0.0444	0.0420	0.0435
RESPY4	0.0145	0.5139	-0.1034	-0.1604	-0.0667
G2	65.657 (0.592)	67.862 (0.516)	76.549 (0.249)	138.129 (1.569e-06)	74.90118 (0.293)
$\chi2$	68.829 (0.483)	79.627 (0.179)	77.389 (0.229)	202.054 (5.551e-15)	76.64593 (0.247)
W2	28.503 (1)	36.455 (0.9996)	39.14861 (0.9986)	76.35734 (0.2541)	32.90313 (0.999)
adj. resd. > 2	3	3	6	7	3

Remark: Y12....Y44: the main effects association terms, Y1score:Y2score is a linear-by-linear association term, α_1= CUT1, α_2=CUT2, RESPY2 is the value of the second response in the marginal model, G2 is the likelihood ratio statistic, $\chi2$ is Pearson's Score statistic, W2is generalized Wald statistic, (*) is the p-value corresponding each test statistic and adj. resd are the number of adjusted residuals which are greater than 2 in each case.

Table 6: Estimates of the AM model for Scenario2 with MAR mechanism using: CC, mode imputation, LOCF, KNNI and MI:

	CC	Mode imputation	LOCF	KNNI	MI
Intercept	-5.2937	-5.1655	-4.7715	-5.9312	-5.2632
Y12	-4.2777	-4.3448	-3.7108	-4.7523	-3.9653
Y13	-8.5653	-8.4365	-7.0132	-9.8770	-7.9749
Y22	-4.6150	-4.9919	-3.7290	-5.3434	-4.4352
Y23	-9.5333	-9.8960	-7.2229	-11.7144	-9.3466
Y32	-4.1929	-3.8009	-3.9311	-4.1040	-4.3706
Y33	-8.2585	-7.1186	-7.5035	-8.0530	-9.1134
Y42	-4.4336	-5.0958	-4.0984	-5.1723	-4.2351
Y43	-8.9556	-10.2177	-8.0274	-11.2510	-8.5387
Y1score:Y2score	0.1283	0.2123	-0.1831	0.2165	0.1700
Y1score:Y3score	0.9712	1.3583	1.3270	0.9115	1.0237
Y1score:Y4score	1.0512	1.0541	0.5225	1.2673	0.7639
Y2score:Y3score	1.0983	0.8567	1.5143	1.1598	1.1041
Y2score:Y4score	1.1905	1.6952	1.4506	1.6176	1.0963
Y3score:Y4score	-0.1440	-0.4763	-0.0516	-0.1754	0.0803
CUT1	0.2443	0.3253	0.3042	0.3267	0.2748
CUT2	0.9506	1.0127	1.0495	1.0146	1.0701
RESPY2	-0.0060	0.4832	-0.1265	-0.0886	-0.0359
RESPY3	0.0835	0.0023	0.0001	-0.0013	0.0007
RESPY4	0.0836	0.5602	-0.0640	-0.0152	0.1402
G2	66.6485 (0.5579)	69.75301 (0.45)	58.4175 (0.814)	94.2958 (0.02327)	64.89653 (0.6177)
χ^2	73.6797 (0.3277)	82.06146 (0.135)	65.121 (0.6101)	132.355 (6.996e-06)	62.90366 (0.6836)
W2	27.89877 (1)	27.5894 (1)	29.310 (1)	56.1732 (0.8666)	27.07242 (1)
adj. resd. > 2	5	5	5	9	5

Remark: Y12....Y44: the main effects association terms, Y1score:Y2score is a linear-by-linear association term, α_1= CUT1, α_2=CUT2, RESPY2 is the value of the second response in the marginal model, G2 is the likelihood ratio statistic, χ^2 is Pearson's Score statistic, W2is generalized Wald statistic, (*) is the p-value corresponding each test statistic and adj. resd are the number of adjusted residuals which are greater than 2 in each case.

5-3 NMAR Simulation Results

Table 7 and Table 8 show the simulation results of scenario 1 and scenario 2 respectively, for estimating the AM model with NMAR missing mechanism and low (10%) missing rate using: CC, mode imputation, LOCF, KNNI and MI.

Thus, after handling MS in the AM model and depending on the goodness-of-fit test statistics and the number of adjusted residuals greater than 2 as evaluation criteria, the best method for handling MS in scenario 1 (Table 7) is LOCF while the worst methods for handling MS are KNNI and MI. Also, the best methods for handling MS in scenario 2 (Table 8) are the LOCF and MI while the worst method for handling MS in scenario 2 is KNNI method.

Table 9 and Table 10 show the simulation results of scenario 1 and scenario 2 respectively, for estimating the AM model with NMAR missing mechanism and high (50%) missing rate using: CC, mode imputation, LOCF, KNNI and MI.

Thus, after handling MS in the AM model and depending on the goodness-of-fit test statistics and the number of adjusted residuals greater than 2 as evaluation criteria, the best method for handling MS in scenario 1 (Table 9) is mode imputation method, while the worst method for handling MS is KNNI. While for scenario 2 (Table 10), the best method for handling MS in scenario 2 is mode imputation method, while the worst method for handling MS is KNNI.

Table 7: Estimates of the AM model for Scenario1 with NMAR and 10% missing rate, using: CC, mode imputation, LOCF, KNNI and MI:

	CC	Mode imputation	LOCF	KNNI	MI
Intercept	-3.6489	-3.5719	-3.5291	-3.7236	-3.6334
Y12	-3.1297	-3.3810	-3.1138	-3.5418	-3.4216
Y13	-5.1686	-5.7508	-5.1855	-6.1335	-5.8417
Y22	-2.6002	-2.7386	-2.4684	-2.9106	-2.6489
Y23	-3.8406	-4.1414	-3.5966	-4.5467	-3.9731
Y32	-2.8685	-3.0539	-2.9988	-3.1171	-2.9656
Y33	-4.4611	-4.9279	-4.8049	-5.0903	-4.7721
Y42	-3.0211	-2.8911	-2.7703	-3.0603	-2.9075
Y43	-4.8629	-4.5677	-4.3486	-4.9655	-4.5967
Y1score:Y2score	0.2235	0.2713	0.2089	0.2501	0.2449
Y1score:Y3score	0.4987	0.6038	0.5585	0.6522	0.6355
Y1score:Y4score	0.5429	0.4985	0.4859	0.5582	0.4966
Y2score:Y3score	0.2953	0.3019	0.2848	0.3552	0.2929
Y2score:Y4score	0.3775	0.3593	0.3226	0.4457	0.3858
Y3score:Y4score	0.2272	0.2180	0.2195	0.1778	0.1920
CUT1	0.0909	0.3062	0.1951	0.3032	0.2772
CUT2	0.8118	1. 00080.	0.9519	1.0228	1.0104
RESPY2	0.0219	0.0080	0.0005	-0.0257	-0.0859
RESPY3	0.0912	0.0445	0.1480	-0.0025	-0.0458
RESPY4	0.0488	0.5760	-0.0236	-0.0561	-0.0489
G2	80.14095 (0.1691)	82.88376 (0.121)	76.36003 (0.254)	96.49399 (0.0161)	not Ho... 97.99125 (0.01244)
χ^2	84.12979 (0.1038)	81.95467 (0.1365)	77.11963 (0.2351)	91.34245 (0.0372)	92.07333 (0.03323)
W2	35.79652 (0.9997)	35.55548 (0.9997)	37.80 (0.9992)	34.38388 (0.9998)	37.38494 (0.9993)
adj. resd. > 2	6	5	6	8	6

Remark: Y12....Y44: the main effects association terms, Y1score:Y2score is a linearby-linear association term, α_1 = CUT1, α_2 =CUT2, RESPY2 is the value of the second response in the marginal model, G2 is the likelihood ratio statistic, χ_2 is Pearson's Score statistic, W2is generalized Wald statistic, (*) is the p-value corresponding each test statistic and adj. resd are the number of adjusted residuals which are greater than 2 in each case.

Table 8: Estimates of the AM model for Scenario2 with NMAR and 10% missing rate, using: CC, mode imputation, LOCF, KNNI and MI:

	CC	Mode imputation	LOCF	KNNI	MI
Intercept	-5.4603	-4.5803	-4.2495	-5.1149	-5.3995
Y12	-4.3888	-4.3834	-4.0624	-4.7767	-4.5260
Y13	-8.7572	-8.3864	-7.4315	-9.7318	-8.9921
Y22	-4.8194	-4.1060	-3.6493	-4.4809	-4.7401
Y23	-9.8598	-7.5583	-6.3460	-8.8798	-9.4746
Y32	-4.7056	-4.0228	-3.3577	-4.1365	-4.8106
Y33	-9.5902	-7.4299	-5.7067	-7.9465	-9.7029
Y42	-3.8838	-3.8235	-3.6095	-4.0370	-4.3278
Y43	-7.3127	-6.8159.	-6.3227	-7.7692	-8.3747
Y1score:Y2score	0.0578	0.3944	0.2303	0.3634	-0.3196
Y1score:Y3score	1.2539	0.8463	0.6734	1.0442	1.3995
Y1score:Y4score	0.9387	0.8310	0.8744	1.0066	1.2099
Y2score:Y3score	1.2916	0.7923	0.6943	0.8752	1.3742
Y2score:Y4score	1.1049	0.6163	0.5933	0.9355	1.2681
Y3score:Y4score	0.2820	0.1559	0.0410	0.0585	-0.4842
CUT1	0.0240	0.3946	0.2967	0.3380	0.2560
CUT2	0.6635	1.0322	0.9547	1.0032	0.8980
RESPY2	0.0765	0.0483	-0.0353	0.0120	0.1047
RESPY3	0.1313	0.0114	-0.1125	0.0657	0.1251
RESPY4	0.0593	0.0467	-0.0110	-0.0153	0.0800
G2	63.881 (0.6517)	82.2255 (0.132)	73.0243 (0.3473)	89.27925 (0.0508)	73.3176 (0.3385)
χ^2	63.905 (0.6509)	86.0299 (0.08064)	74.0374 (0.3173)	104.3109 (0.00389)	73.6018 (0.33)
W2	26.8181 (1)	37.2079 (0.9994)	34.00594 (0.9999)	42.09443 (0.9956)	33.40531 (0.9999)
adj. resd. > 2	5	5	5	9	5

Remark: Y12....Y44: the main effects association terms, Y1score:Y2score is a linear-by-linear association term, α_1= CUT1, α_2=CUT2, RESPY2 is the value of the second response in the marginal model, G2 is the likelihood ratio statistic, χ^2 is Pearson's Score statistic, W2is generalized Wald statistic, (*) is the p-value corresponding each test statistic and adj. resd are the number of adjusted residuals which are greater than 2 in each case.

Table 9: Estimates of the AM model for Scenario1 with NMAR and 50% missing rate, using: CC, mode imputation, LOCF, KNNI and MI:

	CC	Mode imputation	LOCF	KNNI	MI
Intercept	-4.9383	-3.9663	-4.1669	-4.8670	-4.1623
Y12	-2.2953	-3.4817	-2.2670	-4.1923	-2.7286
Y13	-3.4435	-5.8423	-3.9928	-8.9712	-5.0562
Y22	-2.3533	-2.8365	-1.9571	-2.6842	-2.5514
Y23	-3.5729	-4.2905	-3.1896	-4.9974	-4.5829
Y32	-2.5073	-2.8489	-2.1552	-3.7747	-2.2587
Y33	-3.9000	-4.4975	-3.6872	-7.9034	-3.9697
Y42	-2.5774	-1.1160	-2.0553	-2.8743	-2.0439
Y43	-4.0771	-0.6163	-3.4550	-5.5969	-3.4948
Y1score:Y2score	0.1462	0.3619	0.1226	0.0696	0.3785
Y1score:Y3score	0.2887	0.6829	0.5024	1.2537	0.5909
Y1score:Y4score	0.5961	0.2723	0.4313	0.8433	0.2762
Y2score:Y3score	0.4805	0.3574	0.2988	0.5981	0.2095
Y2score:Y4score	0.3325	0.1948	0.3282	0.5225	0.4803
Y3score:Y4score	0.2080	-0.0279	0.0914	0.0547	0.1480
CUT1	-1.2420	0.3055	-0.5820	-0.0200	-0.2958
CUT2	0.2080	1.0472	0.3890	0.9361	0.6772
RESPY2	0.2931	0.0078	0.2445	-0.0634	0.0870
RESPY3	0.4059	0.0444	0.1701	0.0116	0.0314
RESPY4	0.0796	-1.3501	0.1385	-0.1682	-0.1271
G2	67.075 (0.5432)	88.32753 (0.06)	106.159 (0.0027)	153.71 (2.143e-08)	111.4183 (0.00093)
χ^2	70.015 (0.4433)	86.33004 (0.08)	94.6119 (0.0221)	160.58 (2.903e-09)	116.6429 (0.00029)
W2	22.19 (1)	37.62378 (0.9993)	41.83862 (0.996)	57.67014 (0.832)	56.86323 (0.8516)
adj. resd. > 2	7	7	8	13	6

Remark: Y12....Y44: the main effects association terms, Y1score:Y2score is a linear-by-linear association term, α_1 = CUT1, α_2 =CUT2, RESPY2 is the value of the second response in the marginal model, G2 is the likelihood ratio statistic, χ^2 is Pearson's Score statistic, W2is generalized Wald statistic, (*) is the p-value corresponding each test statistic and adj. resd are the number of adjusted residuals which are greater than 2 in each case.

Table 10: Estimates of the AM model for Scenario2 with NMAR mechanism and 50% missing rate using CC, mode imputation , LOCF , KNNI and MI:

	CC	Mode imputation	LOCF	KNNI	MI
Intercept	-5.8158	-3.9750	-3.7355	-4.8911	-5.1016
Y12	-4.5855	-4.3879	-4.3084	-4.5868	-4.4516
Y13	-8.4131	-6.8677	-6.7690	-7.9530	-7.8860
Y22	-5.7768	-4.5656	-3.7920	-5.5885	-5.0794
Y23	-11.6064	-7.2138	-5.4601	-10.6612	-9.5393
Y32	-6.1952	-4.3729	-3.5804	-4.7090	-5.2306
Y33	-12.8159	-6.8554	-5.0359	-8.2949	-9.9473
Y42	-4.2999	-3.9712	-3.8260	-4.0766	-4.3682
Y43	-7.4018	-5.7748	-5.5809	-6.6220	-7.5502
Y1score:Y2score	-0.0765	0.0944	0.1363	0.2988	-0.4043
Y1score:Y3score	1.5518	0.8517	0.7101	1.0296	1.3073
Y1score:Y4score	0.7493	0.7350	0.7644	0.6844	1.1373
Y2score:Y3score	1.6449	0.8515	0.5825	1.1411	1.5590
Y2score:Y4score	1.2376	0.7252	0.5950	1.1819	1.1917
Y3score:Y4score	-0.1539	-0.0725	-0.0613	-0.2563	0.5439
CUT1	0.1778	0.5971	0.4751	0.4075	0.3231
CUT2	0.5988	0.9967	0.9080	0.8278	0.7943
RESPY2	0.1678	0.1325	-0.0757	0.0707	0.1605
RESPY3	0.1644	0.0695	-0.1320	0.1297	0.1896
RESPY4	0.0714	-0.0014	0.0122	-0.0051	0.0995
G2	48.3409 (0.9722)	73.538 (0.332)	69.888 (0.4475)	82.83085 (0.1225)	75.62781 (0.2731)
χ^2	60.578 (0.7552)	81.975 (0.136)	82.03597 (0.1351)	101.1721 (0.0070))	87.32212 (0.06741)
W2	20.4423 (1)	29.4139 (1)	35.18933 (0.9998)	36.29499 (0.9996)	34.48499 (0.9998)
adj. resd. > 2	6	7	6	8	7

Remark: Y12....Y44: the main effects association terms, Y1score:Y2score is a linear-by-linear association term, α_1 = CUT1, α_2 =CUT2, RESPY2 is the value of the second response in the marginal model, G2 is the likelihood ratio statistic, χ_2 is Pearson's Score statistic, W2is generalized Wald statistic, (*) is the p-value corresponding each test statistic and adj. resd are the number of adjusted residuals which are greater than 2 in each case.

6- Conclusions and Recommendations

In this paper, a longitudinal study was considered with four points in time each with three levels, two levels of correlation structures and a total of 200 subjects. Two possible missing rates and five methods for handling MS were used to detect the effect of the methods for handling MS on the AM model. In addition, three missing mechanisms were considered (that is, MCAR, MAR and NMAR). Based on the simulation results, we have reached the following important conclusions as displayed in table 11:

- There is no one single method for handling MS, which is the best under all situations.
- Also, it should be noted that the CC method concludes good imputation method with condition of small missing percentage and large sample size to improve its disadvantages. As Allison (2001) refers, CC method "is not a bad method for handling missing data". Thus, either with low or high missing rate and either with low or high correlation structure the CC method for handling MS leads to a significant AM model.
- For MCAR with low missing rate, either with low or high correlation structure, LOCF was the best method for handling MS in the AM model, while KNNI was the worst method for handling MS in the AM model. Also, for MCAR with high missing rate and low correlation structure, the mode imputation method was the best one, while MI was the worst method. But for MCAR with high missing rate and high correlation structure, none of the methods leads to a significant AM model, where all the methods lead to non significant AM model.
- For the MAR mechanism, the simulation results revealed that MI is the best method regardless of the missing rate and strength of the correlation structure between the points in time. Thus, for low or high correlation structure ,MI was the best method and KNNI was the worst method for handling MS in the AM model
- For NMAR, with low missing rate and low correlation structure, LOCF was the best method for handling MS in the AM model, while KNNI and MI were the worst methods for handling MS in the AM model. Also, for NMAR with low missing rate and high correlation structure, LOCF and MI were the best methods, while KNNI was the worst method. But for NMAR with high

missing rate and either low or high correlation structure, mode imputation was the best method and KNNI was the worst method.

- It should be noted that:

In 40% of the cases, LOCF was the best method,

In 30% of the cases, mode imputation was the best method,

In 20% of the cases, MI was the best method,

In 10% of the cases, none of the methods leads to a significant AM model.

Also, in 80% of the cases, KNNI was the worst method.

Table 11: Results of the simulation study for each scenario:

Scenario	Missing Mechanism	Missing Rate	Best Method	Worst Method
1 (N=200, J=3, ρ =0.2)	MCAR	10%	LOCF	KNNI
	MCAR	50%	mode imp.	MI
	MAR	---	MI	KNNI
	NMAR	10%	LOCF	KNNI, MI
	NMAR	50%	mode imp.	KNNI
2 (N=200, J=3, ρ=0.6)	MCAR	10%	LOCF	KNNI
	MCAR	50%	………	……….
	MAR	----	MI	KNNI
	NMAR	10%	LOCF, MI	KNNI
	NMAR	50%	mode imp.	KNNI

7-REFERENCES

Allison, P. (2002). Missing Data, Sage Publications, *Thousand Oaks*.

Al-Zahrani, H. (2018). Missing Data Analysis for Binary Multivariate Longitudinal Data through a simulation Study. *Biometrics and Biostatistics International Journal*, 7(2), 103-113

Baraldi, A. & Enders, C. (2010). An Introduction to Modern Missing Data Analyses. *Journal of School Psychology, 48, 5-37*.

Bori, M. (2013). *Dealing with Missing Data: Key Assumptions and Methods for Applied Analysis.*

Deng, Y.; Chang, C.; Ido, C. and Long, Q. (2016). Multiple Imputation for General Missing Data Patterns in the Presence of High-Dimensional Data. *Scientific Reports, 6.*

Garg, P. (2013). Robustness of Multiple Imputation under Missing At Random (MAR) Mechanism: A simulation Study. *Theses and Dissertations, 35, 1-206.*

Horton, N. and Kleinman, K. (2007). Much Ado About Nothing: A Comparison of Missing Data Methods and Software to Fit Incomplete Data Regression Models. *The American Statistician, 61, 79-90*.

Kombo, A.; Mwambi, H. and Molenberghs, G. (2017). Multiple Imputation for Ordinal Longitudinal Data with Nonmonotone Missing Data Patterns. *Journal of Applied Statistics, 44(2), 270-287*

Lang, J. and Agresti, A. (1994). Simultaneously Modeling Joint and Marginal Distributions of Multivariate Categorical Responses, *Journal of the American Statistical Association, 89, 625-632.*

Lang, J.; Mcdonald, J. and Smith, P. (1997). Association-Marginal Modeling of Multivariate Categorical Responses. A Maximum Likelihood Approach for Large, Sparse Tables. *University of Iowa: Technical Report. 263.*

Lang, J. and Eliason, S. (1997). Applications of Association-Marginal Models to the Study of Social Mobility. *Sociological Methods and Research, 26, 183-212.*

Nakai, M.; Chen, D.; Nishimura, K. and Miyamoto, Y. (2014). Comparative Study of Four Methods in Missing Value Imputations Under Missing Completely at Random Mechanism. *Open Journal of Statistics, 4, 27-37.*

Nooraee, N.; Molenbergh, G.; Ormel, J. and Heuvel, E. (2018). Strategies for Handling Missing Data in Longitudinal Studies with Questionnaires.

Rindskopf, D. (1992). A General Approach to Categorical Data Analysis with Missing Data, Using Generalized Linear Models with Composite Links. *Psychometrika, 57, 29-42.*

Rithy, D. (2016). *Simulation of Imputation Effects under Different Assumptions*.

Rubin, D. (1976). Inference and Missing Data. *Biometrika, 63, 581-590.*

Rubin, D. (1987). *Multiple Imputation for Nonresponse in Surveys.* New York: John Wiley and Sons.

Schlomer, G.; Bauman, S. and Card, N. (2010). Best Practices for Missing Data Management in Counseling Psychology. *Journal of Counseling Psychology, 57, 1-10.*

Touloumis, A. (2016). Simulating Correlated Binary and Multinomial Responses under Marginal Model Specification: The SimCorMultRes Package. *The R Journal, 8 (2), 79 – 91*

Touloumis, A. (2018). *Simulating Correlated Binary and Multinomial Responses with SimCorMultRes.*

Zhu, X. (2014). Comparison of Four Methods for Handling Missing Data in Longitudinal Data Analysis Through a Simulation Study. *Open Journal of Statistics, 4, 933-944.*

Suggested Citation according to APA Style

El-Zayat, M.; Mohamed, E.; Halawa, A. and El-Attar, L. (2019). Handling missing data in the Association-Marginal. *Journal of the Faculty of Comemrce for Scientific Research*, Faculty of Commerce, Alexandria University, 56(3), 187 – 212.