Biomedical Statistics and Informatics
2023; 8(2): 22-30
http://www.sciencepublishinggroup.com/j/bsi
doi: 10.11648/j.bsi.20230802.11
ISSN: 2578-871X (Print); ISSN: 2578-8728 (Online)

A Gaussian Copula Regression Approach for Modelling Repeated Data in Medical Research

Reka Karuppusami[1, *], Gomathi Sudhakar[1], Juliya Pearl Joseph Johnson[2], Ramamani Mariappan[3], Jansi Rani[1], Belavendra Antonisamy[1], Prasanna S. Premkumar[1]

[1]Department of Biostatistics, Christian Medical College, Vellore, India

[2]Departments of Anaesthesia, Christian Medical College, Vellore, India

[3]Department of Neuroanaesthesia, Christian Medical College, Vellore, India

Email address:

reachreka28@gmail.com (Reka Karuppusami), gomathisudhakar1998@gmail.com (Gomathi Sudhakar), juliya_pearl@yahoo.com (Juliya Pearl Joseph Johnson), ramamani@cmcvellore.ac.in (Ramamani Mariappan), janrandai@gmail.com (Jansi Rani), b.antonisamy@gmail.com (Belavendra Antonisamy), prasanna_samuel@hotmail.com (Prasanna S. Premkumar)

[*]Corresponding author

To cite this article:

Reka Karuppusami, Gomathi Sudhakar, Juliya Pearl Joseph Johnson, Ramamani Mariappan, Jansi Rani, Belavendra Antonisamy, Prasanna S. Premkumar. A Gaussian Copula Regression Approach for Modelling Repeated Data in Medical Research. *Biomedical Statistics and Informatics*. Vol. 8, No. 2, 2023, pp. 22-30. doi: 10.11648/j.bsi.20230802.11

Received: June 27, 2023; **Accepted**: July 19, 2023; **Published**: July 31, 2023

Abstract: In repeated measures data, the observations tend to be correlated within each subject, and such data are often analyzed using Generalized Estimating Equations (GEE), which are robust to assumptions that many methods hold. The main limitation of GEE is that its method of estimation is quasi-likelihood. The recent framework of the copula is very popular for handling repeated data. The maximum likelihood-based analysis for repeated data can be obtained using Gaussian copula regression. The purpose of this study is to show the handling and analysis of the repeated data using the Gaussian copula regression approach and compare the findings with GEE. The prospective, double-blinded, randomized controlled trial data for this study was obtained from the Department of Anesthesia, Christian Medical College, and Vellore. ASA I and II patients were randomized into three groups. Hemodynamic parameters were obtained for 88 patients at thirteen-time points. The outcome of interest was mean arterial pressure. Both GEE and Gaussian copula regression were compared assuming four different correlation structures. The optimal correlation structures were selected with the Akaike Information Criterion (AIC) and Correlation Information Criterion (CIC) goodness of fit criteria according to the method of estimation of Gaussian copula regression and GEE, respectively. The correlation structures, unstructured and autoregressive, were found to be optimal for Gaussian copula regression and GEE based on AIC and CIC criteria values respectively. A comparison between the estimated values of the selected models showed no major differences. Gaussian copula regression found that intrathecal morphine has a significant reduction in MAP over time, this significance is considered important as the study uses randomized controlled trial data. Both methods have almost similar results, but Gaussian copula regression provides better results by identifying significant findings associated with the outcome using maximum likelihood estimation that GEE fails to identify using quasi-likelihood estimation.

Keywords: Correlation Structures, Gaussian Copula Regression, Generalized Estimating Equations, Repeated Data

1. Introduction

Repeated data arise frequently in biomedical and many other research fields. In longitudinal studies, the correlation usually occurs when data are collected sequentially from the same individual over time. Statistical procedures that fail account for the correlation between repeated data points are likely to yield invalid conclusions due to parameter estimates may be inconsistent with wrong standard error estimates. [1] Over the years, there are many statistical methods available

for handling and analyzing repeated data such like Repeated Measures ANOVA (RM-ANOVA), Mixed Effects Model (MEM) and Generalized Estimating Equations (GEE). These statistical procedures will be used under some specific conditions and properties [2-4].

Repeated measure studies are always important and offer the chance to observe individual patterns of change over time. The generalized estimating equation method is widely used in practice for analyzing repeated measurement data. [2, 5, 6] The generalized estimating equations have started to replace the traditional methods such as Repeated Measures Analysis Of Variance as the older method is not flexible enough to accommodate all of the features of repeated measure designs [7, 8]. But in GEE, the model selection is difficult due to the lack of an absolute goodness of fit test to support the selection of the best model among several models. [9-11]

The recent growing framework of the copula approach is becoming popular and used to handle repeated measure data. [12] The word copula means 'a bond, link or tie', and was first employed in a statistical or mathematical sense by Abe Sklar in 1959. [12, 13] The copula method is relatively different and is applied in various fields such as spatial medical, financial, and air pollution. [14-20] A Copula is a function that permits us to combine univariate distributions to get a joint distribution with a particular dependence structure. [21–23] The main advantages of copula methodology in modeling repeated data are pointed out here: (1) allowance to model both linear and non-linear dependence, (2) capability of modeling extreme endpoints, and (3) arbitrary choice of a marginal distribution. [21, 22, 24]

As an alternative to GEE, a Gaussian copula regression approach under the copula framework is proposed for modeling repeated data. [22] The Gaussian copula regression model is very flexible and is used to analyze repeated data even with missing values and when the sample size is small. The Gaussian copula estimation procedure for the estimation of regression parameters in repeated response data using various working correlation structure is compared with the generalized estimating equations. The objective of this study is to show the method of handling repeated data using the Gaussian copula regression approach with real data from a published randomized controlled trial. [25]

2. Methods

2.1. Data

The data used for this study is from a prospective, double-blinded, randomized control trial that was conducted over two years between November 2016 and November 2018 in ASA I and II patients aged between 18-60 years with normal renal function who underwent substitutional urethroplasty with buccal mucosal graft. Patients were randomized into three groups, group-A received systemic morphine (0.1mg/kg), group-B received epidural morphine (3mg), and group-C received intrathecal morphine (150µg). Repeated measurements of hemodynamic parameters (systolic blood pressure, diastolic blood pressure, and mean arterial pressure) were collected for each patient at thirteen various time points. The mean arterial pressure was considered an outcome for analysis. This study was approved by the Institutional Review Board (IRB Number-10285; dated 21-09-2016) and ethics committee of Christian Medical College (CMC) and Hospital, Vellore.

2.2. Study Variables

Baseline demographic variables, associated co-morbidities, and other study characteristics of interest included were age (years), BMI (kg/m^2), previous surgery (yes or no), allergies (yes or no), alcohol consumption (yes or no), smoker (yes or no), diabetes mellitus (yes or no), hypertension (yes or no), and mean arterial pressure (MAP).

2.3. Statistical Analysis

The descriptive statistics were reported as Mean ± SD for continuous and number (percentage) for categorical data. The Pearson Chi-square test and Fisher's exact test (less cell frequency) were used to find the association between categorical variables. A One-way analysis of variance was performed to find the difference between groups on the continuous data. Further, the only significant predictor variable, hypertension was included in the model. A line plot was used to visualize a trend in MAP between three groups over time. Two analytical methods, GEE and Gaussian copula regression were compared. Unadjusted and adjusted estimates with 95% confidence intervals and p-values were reported for both GEE and Gaussian copula regression. Gaussian copula regression and GEE were performed using the *gcmr* and *geepack* packages in R software version 4.1.0. All statistical tests were two-sided at α=0.05 level of significance.

2.4. Generalized Estimating Equations

Many simple approaches exist for handling repeated data and analysis, but the limit is the inability to include covariates. GEE is a method to fit a marginal model for longitudinal data analysis, and it has been widely used in biomedical and clinical trial research. [26, 27] The GEE is based on a quasi-likelihood function and provides robust estimates of the parameter. [28] In Generalized Estimating Equations, the within-subject correlation structure is treated as a 'nuisance' variable (i.e., as a covariate). [29] GEE has many features 1) The variance-covariance matrix of responses is treated as nuisance parameters in GEE and therefore this model fitting turns out to be easier than another approach of mixed-effect models. 2) Specifically, if the treatment effect is of primary interest to the research study, the GEE approach is preferred. 3) GEE relaxes the distribution assumption and only needs the correct specification of marginal mean and variance as well as the link function which connects the covariates of interest and marginal means. [30-34]

However, generalized estimating equations are still

controversial in many aspects, like some issues with inconsistent estimation of within-subject correlation coefficients under a misspecified "working" correlation structure. [35] The estimation of the correlation coefficients using the moment-based method is not efficient. Such limitations lead investigators or researchers to actively work on this area to develop novel methods. Many alternative methods for estimating the correlation coefficients have been offered; among those, one method was based on the "Gaussian" estimation approach. [20, 21]

2.5. Gaussian Copula Regression

There are different types of copulas used in modeling. The Gaussian copula under the elliptical copula family has been successfully employed in several complex applications arising, for example, in repeated data analysis. For longitudinal data applications, elliptical copulas are more useful than Archimedean copulas. [36] Elliptical copulas have a correlation structure described with a correlation matrix that can handle the time-series behavior of repeated data. Copula function has some effect on the shape of the joint distribution, so an appropriate and reasonable copula function should be selected because the properties exhibited by different copula functions will vary. [19]

The regression model is expressed as

$$Y_i = g(x_i, \epsilon_i; \lambda), i = 1,2,3, \dots, n. \quad (1)$$

Where g (.) is a suitable function of the repressors x_i and of an unobserved stochastic variableϵ_i, denoted as the error term. The regression model (1) is known up to a vector of parameters λ.

The useful choice of g(.) is

$$Y_{i = F_i^{-1}\{\phi[\epsilon_i]; \lambda\}}, i = 1,2,3, \dots., n. \quad (2)$$

$\in_i$ is a standard normal variable and $F_i(.; \lambda) = F(.|x_i; \lambda)$ and ϕ(.) are the cumulative distribution functions of Y_i given x_i and of a standard normal variate respectively. [22]

The Gaussian copula was first proposed for modeling selectivity in the context of continuous but non-normal distributions. Gaussian copula regression provides a general framework for modeling outcome variables that may belong to any distribution family. It is flexible enough to allow for both positive and negative dependencies. The dependence observed can be expressed with a convenient working correlation structure like autoregressive or exchangeable. [21, 22, 24] Inference for continuous responses is carried out through the likelihood approach, for non-continuous responses, numerical approximations are used.

3. Results

A total of 93 male patients were randomized for the study into three groups, with 31 patients in Group 1, 31 patients in Group 2, and 31 patients in Group 3. Of these, 2 patients were excluded in Groups I and II, and 1 in Group 3, respectively, after randomization. [7] For analysis, 88 male patients were included aged between 18 and 67 years. The mean age of the patient is 39 years (SD 12. 2). Table 1 describes the summary statistics of baseline demographic and other characteristics of study patients who were allocated into three groups receiving systemic morphine, epidural morphine, and intrathecal morphine. We found that only hypertension was significantly associated with the group (P = 0.049).

Table 2 describes the mean arterial pressure that was compared between groups at each time point. We found a statistically significant difference between groups on MAP at 1.5 hours. Figure 1 shows the correlation between sets of time points on the MAP. The correlogram allows that to see which pairs have the highest correlation. Figure 2 explains graphically how the changes vary over time at thirteen-time points between the three groups. From there, it was observed that until 30 minutes, there were no major changes between groups. After 30 minutes, there is a drastic drop in the mean arterial pressure for the patients in the intrathecal morphine group compared to systemic morphine and epidural morphine.

GEE analyses were done by assuming common four correlation structures, namely independence, autoregressive, exchangeable, and unstructured. The time at which repeated measurements have been taken for each patient and the treatment groups of the patients were considered factors (time and group respectively). In Table 3, unadjusted GEE analysis was carried out for various correlation structures considering the factors time, group, and an interaction term of groups over time. There is a significant change over time in mean arterial pressure, and intrathecal morphine was shown to have a reduction in mean arterial pressure when compared to systemic morphine, assuming independence and exchangeable correlation structures. Since the presence of hypertension among the patients showed a significant association between groups from the baseline analysis, hypertension was considered in the adjusted GEE model along with time, groups, and interaction between groups and time which is given in Table 4. Similar to the results of the GEE unadjusted model, in the GEE adjusted model, there was also a significant change in the mean arterial pressure over time and hypertension.

The unadjusted and adjusted Gaussian copula regressions were performed in Tables 5 and 6 respectively. There was a significant reduction in mean arterial pressure over time. In addition, in the unstructured correlation structure, over time the intrathecal morphine group had a significant reduction in the mean arterial pressure compared to the systemic morphine group. The Gaussian copula was an efficient statistical method to capture this finding in unadjusted and adjusted models, but GEE failed to identify it.

The goodness of fit comparison is given in Table 7. The AIC and CIC criteria measures were used to select the optimal correlation structure under Gaussian copula regression and GEE respectively. The goodness of fit criteria may not select the same correlation structure as the method of assessing each criterion differs. An unstructured correlation structure from Gaussian copula regression and an autoregressive correlation structure from GEE were found to

be optimal models as they have smaller values. Overall, there was no major difference in the performance of the two methods.

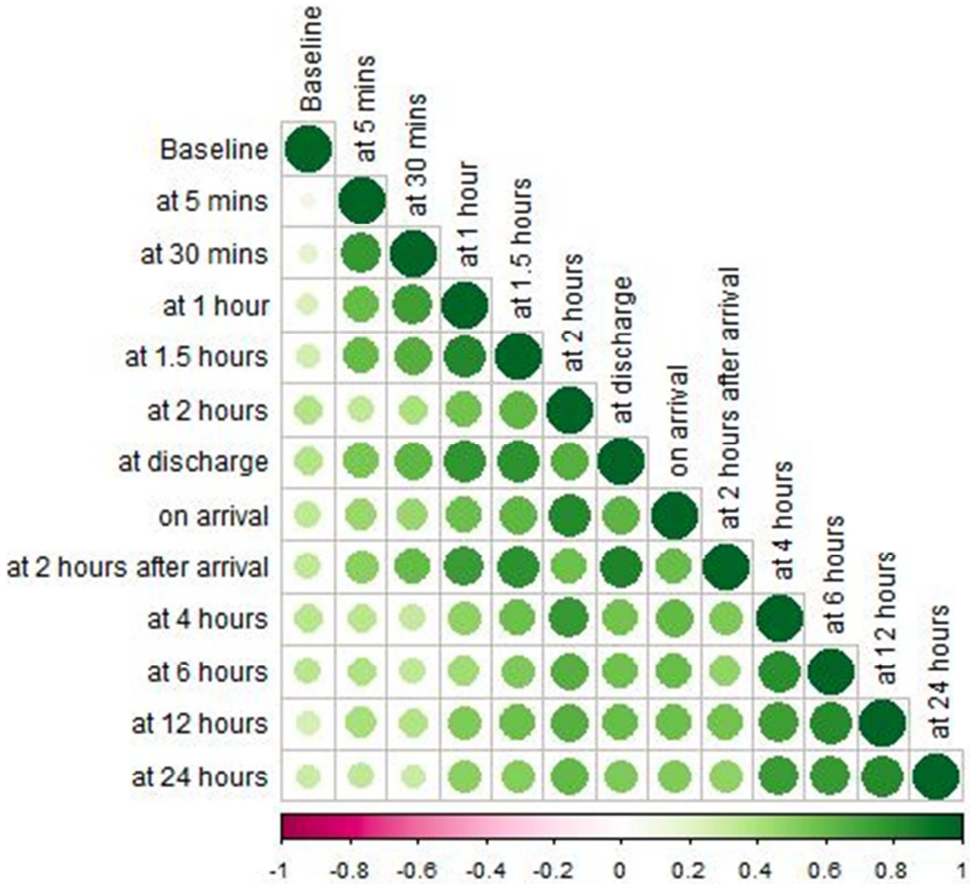

Figure 1. *Correlation matrix plot (Correlogram). Positive correlations are displayed in a green scale while negative correlations are displayed in a pink scale. Correlation between each time points of mean arterial pressure (MAP) was given.*

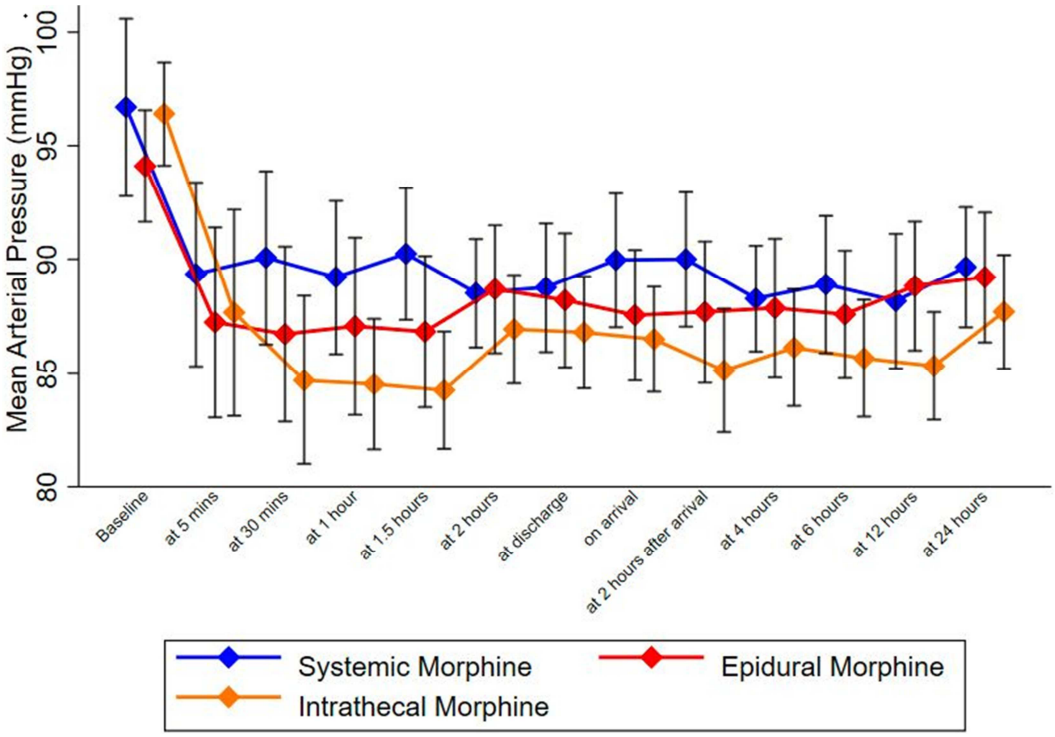

Figure 2. *Line plot showing mean arterial pressure change over time between three groups with 95% confidence interval.*

Table 1. *Baseline demographic and other characteristics of study patients (n=88).*

Variable	Systemic morphine (0.1mg/kg) (n=29)	Epidural morphine (3mg) (n=29)	Intrathecal morphine (150µg) (n=30)	p-value
Age (years)*	38.17±11.90	40.69±12.93	39.47±12.12	0.740
BMI (kg/m^2)*	24.82±3.31	24.72±3.83	25.63±4.28	0.604
Previous surgery				
Yes	11 (37.9)	20 (69.0)	17 (56.7)	0.620
No	18 (62.1)	9 (31.0)	13 (43.3)	
Allergies				
Yes	1 (3.4)	3 (10.3)	0 (0.0)	0.123
No	28 (96.6)	26 (89.7)	30 (100.0)	
Alcohol consumption				
Yes	4 (13.8)	7 (24.1)	7 (23.3)	0.553
No	25 (86.2)	22 (75.9)	23 (76.7)	
Smoker				
Yes	4 (13.8)	8 (27.6)	8 (26.7)	0.373
No	25 (86.2)	21 (72.4)	22 (73.3)	
Diabetes Mellitus				
Yes	4 (13.8)	6 (20.7)	1 (3.3)	0.121
No	25 (86.2)	23 (79.3)	29 (96.7)	
Hypertension				
Yes	7 (24.1)	5 (17.2)	1 (3.3)	0.049
No	22 (75.9)	24 (82.8)	29 (96.7)	

BMI: Body Mass Index; SD: Standard Deviation.
Values are presented as number (percentage) and p-value is obtained from the Chi-square test and Fisher's exact test (less cell count).
*Values are presented as Mean ± SD and p-value is obtained from One-way ANOVA.

Table 2. *Summary statistics for hemodynamic parameter mean arterial pressure.*

Time points	Systemic morphine (0.1mg/kg) (n=29)	Epidural morphine (3mg) (n=29)	Intrathecal morphine (150µg) (n=30)	p-value
	Mean±SD	Mean±SD	Mean±SD	
Baseline	96.69±10.70	94.10±6.72	96.40±6.35	0.419
at 5 mints.	89.34±11.06	87.24±11.42	87.67±12.61	0.771
at 30 mints	90.07±10.42	86.72±10.50	84.70±10.32	0.143
at 1 hour	89.21±9.22	87.07±10.65	84.53±7.96	0.163
at 1.5 hours	90.24±7.94	86.83±9.07	84.27±7.17	0.021
at 2 hours	88.52±6.51	88.69±7.69	86.93±6.60	0.565
at discharge	88.76±7.74	88.21±8.04	86.80±6.81	0.590
on arrival	89.97±8.06	87.55±7.82	86.50±6.39	0.195
at 2 hours after arrival	90.00±8.11	87.69±8.48	85.13±7.54	0.073
at 4 hours	88.28±6.35	87.86±8.33	86.13±7.12	0.492
at 6 hours	88.90±8.26	87.59±7.65	85.67±7.13	0.273
at 12 hours	88.17±8.08	88.83±7.76	85.33±6.56	0.168
at 24 hours	89.66±7.22	89.21±7.82	87.70±6.92	0.563

*SD: Standard Deviation.

Table 3. *Unadjusted parameter estimates using a generalized estimating equation for mean arterial pressure.*

Variable	Correlation structure							
	Independence		Autoregressive [AR (1)]		Exchangeable		Unstructured	
	β (95% CI)	p-value	β (95% CI)	p-value	β (95% CI)	p-value	β (95% CI)	p-value
Time	-0.22 (-0.37,-0.08)	0.001	-0.51 (-0.66, -0.36)	0.001	-0.20 (-0.31,-0.08)	0.001	-0.31 (-0.44, -0.19)	0.001
Group								
Epidural morphine	-1.55 (-4.85, 1.74)	0.356	-1.54 (-4.61, 1.54)	0.328	-1.45 (-4.80, 1.64)	0.356	-1.87 (-9.01, 5.26)	0.607
Intrathecal morphine	-3.08 (-6.05,-0.10)	0.042	-2.16 (-5.03, 0.70)	0.140	-3.10 (-6.01,-0.16)	0.040	-1.33 (-7.97, 5.32)	0.696
Systemic morphine	Ref.		Ref.		Ref.		Ref.	
Group * Time								
Epidural morphine * Time	0.22 (-0.13, 0.56)	0.223	0.18 (-0.19, 0.56)	0.340	0.28 (-0.12, 0.56)	0.223	0.18 (-0.12, 0.49)	0.249
Intrathecal morphine * Time	0.003 (-0.34, 0.35)	0.983	-0.11 (-0.47, 0.25)	0.542	0.007 (-0.31, 0.38)	0.983	-0.04 (-0.37, 0.28)	0.795
Systemic morphine * Time	Ref.		Ref.		Ref.		Ref.	

*95% CI: 95% Confidence Interval

Table 4. *Adjusted parameter estimates using a generalized estimating equation for mean arterial pressure.*

Variable	Correlation structure							
	Independence		Autoregressive [AR (1)]		Exchangeable		Unstructured	
	β (95% CI)	p-value	β (95% CI)	p-value	β (95% CI)	p-value	β (95% CI)	p-value
Hypertension								
Yes	5.12 (1.55, 8.68)	0.004	5.82 (2.70, 8.94)	0.0002	5.12 (1.55, 8.68)	0.004	6.81 (3.55, 10.1)	0.001
No	Ref.		Ref.		Ref.		Ref.	
Time	-0.30 (-0.55, -0.04)	0.022	-0.53 (-0.82, -0.23)	0.0005	-0.30 (-0.55, -0.04)	0.022	-0.35 (-0.60, -0.10)	0.006
Group								
Epidural morphine	-2.72 (-6.99, 1.54)	0.210	-2.44 (-6.28, 1.41)	0.213	-2.72 (-6.99, 1.54)	0.210	-2.19 (-5.93, 1.54)	0.249
Intrathecal morphine	-2.04 (-6.12, 2.03)	0.326	-0.23 (-3.91, 3.44)	0.900	-2.04 (-6.12, 2.03)	0.326	-0.73 (-4.17, 2.70)	0.675
Systemic morphine	Ref.		Ref.		Ref.		Ref.	
Group * Time								
Epidural morphine * Time	0.21 (-0.13, 0.56)	0.222	0.18 (-0.19, 0.56)	0.333	0.21 (-0.13, 0.56)	0.222	0.16 (-0.14, 0.48)	0.291
Intrathecal morphine * Time	0.003 (-0.34, 0.35)	0.983	-0.11 (-0.47, 0.25)	0.549	0.003 (-0.34, 0.35)	0.983	-0.06 (-0.40, 0.26)	0.682
Systemic morphine * Time	Ref.		Ref.		Ref.		Ref.	

*95% CI: 95% Confidence Interval

Table 5. *Unadjusted Gaussian copula regression estimates for mean arterial pressure.*

Variable	Correlation structure							
	Independence		Autoregressive [AR (1)]		Exchangeable		Unstructured	
	β (95% CI)	p-value	β (95% CI)	p-value	β (95% CI)	p-value	β (95% CI)	p-value
Time	-0.22 (-0.36,-0.09)	0.001	-0.39 (-0.60,-0.19)	0.001	-0.22 (-0.32,-0.13)	0.001	-0.21 (-0.36,-0.05)	0.016
Group								
Epidural morphine	-1.55 (-2.78,-0.33)	0.355	-1.54 (-3.83,0.74)	0.334	-1.50 (-4.69,1.58)	0.353	-2.46 (-5.83,0.89)	0.151
Intrathecal morphine	-3.07 (-4.29,-1.86)	0.042	-2.68 (-4.95,-0.41)	0.069	-3.09 (-6.19,-6.19)	0.041	-2.63 (-6.01,0.75)	0.117
Systemic morphine	Ref.		Ref.		Ref.		Ref.	
Group * Time								
Epidural morphine * Time	0.21 (-0.10,0.54)	0.223	0.19 (-0.29,0.68)	0.257	0.21 (-0.01,0.45)	0.222	-0.07 (-0.36,0.22)	0.564
Intrathecal morphine * Time	0.003 (-0.31,0.32)	0.983	-0.06 (-0.55,0.41)	0.692	0.003 (-0.22,0.23)	0.983	-0.25 (-0.55,0.04)	0.044
Systemic morphine * Time	Ref.		Ref.		Ref.		Ref.	

*95% CI: 95% Confidence Interval

Table 6. *Adjusted Gaussian copula regression estimates for mean arterial pressure.*

Variable	Correlation structure							
	Independence		Autoregressive [AR (1)]		Exchangeable		Unstructured	
	β (95% CI)	p-value	β (95% CI)	p-value	β (95% CI)	p-value	β (95% CI)	p-value
Hypertension								
Yes	5.11 (3.71,6.52)	0.004	5.42 (2.86,7.97)	0.001	5.11 (1.55,8.67)	0.005	4.21 (-0.80,9.22)	0.050
No	Ref.		Ref.		Ref.		Ref.	
Time	-0.29 (-0.52,-0.07)	0.022	-0.43 (-0.76,-0.09)	0.001	-0.29 (-0.46,-0.13)	0.022	-0.10 (-0.35,0.14)	0.374
Group								
Epidural morphine	-2.72 (-5.24,-0.19)	0.210	-2.56 (-6.51, 1.38)	0.187	-2.72 (-6.16,0.72)	0.212	-1.85 (-6.42,2.70)	0.368
Intrathecal morphine	-2.04 (-4.56,0.48)	0.326	-1.14 (-5.09,2.81)	0.540	-2.04 (-5.52,1.44)	0.326	0.39 (-4.25,5.04)	0.837
Systemic morphine	Ref.		Ref.		Ref.		Ref.	
Group * Time								
Epidural morphine * Time	0.21 (-0.10,0.53)	0.222	0.19 (-0.27,0.67)	0.247	0.21 (-0.01,0.45)	0.223	-0.14 (-0.45,0.17)	0.312
Intrathecal morphine * Time	0.003 (-0.31,0.31)	0.983	-0.06 (-0.53,0.40)	0.711	0.003 (-0.22,0.23)	0.983	-0.37 (-0.70,-0.05)	0.009
Systemic morphine * Time	Ref.		Ref.		Ref.		Ref.	

*95% CI: 95% Confidence Interval

Table 7. *Goodness of fit criteria.*

Correlation structure	Gaussian copula regression AIC (smaller is better)	Generalized Estimating Equation CIC (smaller is better)
Independence	8114	29.3
Autoregressive [AR(1)]	7647	4.67
Exchangeable	7647	10.5
Unstructured	7277	7.76

4. Discussion

The performance of two statistical methods was assessed with application to repeated data from a RCT study with an outcome of mean arterial pressure. In the present study, Gaussian copula regression was shown to be an efficient method to capture the significant findings. Similar finding of Gaussian copula regression via vector generalized linear model (VGLM), which was found to be more efficient than GEE. [37] GEE estimates using autoregressive correlation structure were compared to estimates obtained from Gaussian copula regression with autoregressive correlation structure. [24] Misleading estimates can be avoided when using maximum likelihood estimation by Gaussian copula regression, which is a major strength if the sample size is small. [38]

Many statistical procedures with maximum likelihood estimation can be used to analyze the repeated data, but they may not be as robust as Gaussian copula regression. For instance, linear mixed-effects models are shown to have the impossibility of joint modeling. [23] In this study, we compared all correlation structures between two statistical methods. However, choosing the same optimum correlation structure for two methods differs due to goodness of fit criteria, which was found to be a limitation. [39]

In this study, the Gaussian copula regression identified certain significantly associated variables using maximum likelihood estimation that even the widely used GEE failed to capture using quasi-likelihood estimation. Although comparisons were made between two different methods of estimation, the model showed similar performance comparatively. From this study, we can find that over time, intrathecal morphine plays a more significant role in the reduction of mean arterial pressure post-surgery than systemic morphine.

5. Conclusion

This study focused on analyzing and comparing the findings of GEE, which uses quasi-likelihood estimation, and Gaussian copula regression, which uses maximum likelihood estimation, when applied to repeated data from a randomized controlled trial. GEE and Gaussian copula regression are highly comparable and lead to valid results. In conclusion, both methods showed similar performance. An alternative to GEE that uses quasi-likelihood estimation, Gaussian copula regression, identifies a few more significant variables using the maximum likelihood method of estimation, but GEE fails.

Funding

The study design and procedures were approved by the Institutional Review Board of the Christian Medical College and Hospital, Vellore, India (IRB Number-10285; dated 21-09-2016). The study was conducted in accordance with the Declaration of Helsinki. The informed consent form was obtained from patients. The Institutional research board of Christian Medical College funded the study.

Author Contributions

Conceptualization: RK, BA; Design of study and acquisition of data: RK, JPJJ, and RM; Analysis and interpretation: RK, GS, and JR; Drafting: RK and GS. Approval of versions: RK, RM and PSP. All authors made material contributions to the handling and intellectual content of this article. All authors read and approved the final manuscript.

Conflict of Interest

None

Abbreviations

AIC: Akaike Information Criterion;
CMC: Christian Medical College;
CIC: Correlation Information Criterion;
CI: Confidence Interval;
GEE: Generalized Estimating Equations;
MEM: Mixed Effects Model;
IRB: Institutional Review Board;
MAP: Mean Arterial Pressure;
RCT: Randomized Control Trial;
RM-ANOVA: Repeated Measures ANOVA;
SD: Standard Deviation;
VGLM: Vector Generalized Linear Model.

References

[1] A D. An Introduction to Generalized Linear Models Florida: Chapman & Hall/CRC. 2002.

[2] PJ D, P H, K L, et al. Analysis of Longitudinal Data. Second Edition. Oxford:: Oxford University Press 2002.

[3] Lee U, Garcia TP, Carroll RJ, et al. Analysis of repeated measures data in nutrition research. Front Biosci (Landmark Ed) 2019; 24: 1377–89.

[4] K. Song PX. Correlated Data Analysis: Modeling, Analytics, and Applications. 2007.

[5] Lu B, Preisser JS, Qaqish BF, et al. A comparison of two bias-corrected covariance estimators for generalized estimating equations. Biometrics 2007; 63: 935–41. doi: 10.1111/j.1541-0420.2007.00764.x.

[6] Crowder M. On the Use of a Working Correlation Matrix in Using Generalised Linear Models for Repeated Measures. Biometrika 1995; 82: 407–10. doi: 10.2307/2337417.

[7] Zeger SL, Liang KY. Longitudinal data analysis for discrete and continuous outcomes. Biometrics 1986; 42: 121–30.

[8] Girden ER. ANOVA: Repeated measures: Sage. 1992.

[9] Horton NJ, Bebchuk JD, Jones CL, et al. Goodness-of-fit for GEE: an example with mental health service utilization. Stat Med 1999; 18: 213–22. doi: 10.1002/(sici)1097-0258(19990130)18:2<213::aid-sim999>3.0.co;2-e.

[10] Barnhart HX, Williamson JM. Goodness-of-fit tests for GEE modeling with binary responses. Biometrics 1998; 54: 720–9.

[11] Heagerty PJ, Zeger SL. Marginalized multilevel models and likelihood inference (with comments and a rejoinder by the authors). Statistical Science 2000; 15: 1–26. doi: 10.1214/ss/1009212671.

[12] Shen C, Weissfeld L. A copula model for repeated measurements with non-ignorable non-monotone missing outcome. Stat Med 2006; 25: 2427–40. doi: 10.1002/sim.2355.

[13] Sklar M. Fonctions de repartition an dimensions et leurs marges. Publ inst statist univ Paris. 1959.

[14] Han Z, De Oliveira V. Maximum Likelihood Estimation of Gaussian Copula Models for Geostatistical Count Data. Commun Stat Simul Comput 2020; 49: 1957–81. doi: 10.1080/03610918.2018.1508705.

[15] Ganjali M, Baghfalaki T. A Copula Approach to Joint Modeling of Longitudinal Measurements and Survival Times Using Monte Carlo Expectation-Maximization with Application to AIDS Studies. J Biopharm Stat 2015; 25: 1077–99. doi: 10.1080/10543406.2014.971584.

[16] Suresh K, Taylor JMG, Tsodikov A. A Gaussian copula approach for dynamic prediction of survival with a longitudinal biomarker. Biostatistics 2021; 22: 504–21. doi: 10.1093/biostatistics/kxz049.

[17] Kim JM, Ju H, Jung Y. Copula Approach for Developing a Biomarker Panel for Prediction of Dengue Hemorrhagic Fever. Annals of Data Science 2020; 7: 697–712. doi: 10.1007/s40745-020-00293-x.

[18] Zhang A, Fang J, Hu W, et al. A Latent Gaussian Copula Model for Mixed Data Analysis in Brain Imaging Genetics. IEEE/ACM Trans Comput Biol Bioinform 2021; 18: 1350–60. doi: 10.1109/TCBB.2019.2950904.

[19] Li J, Zhu X, Lee C-F, et al. On the aggregation of credit, market and operational risks. Rev Quant Finan Acc 2015; 44: 161–89. doi: 10.1007/s11156-013-0426-0.

[20] Masseran N. Modeling the Characteristics of Unhealthy Air Pollution Events: A Copula Approach. Int J Environ Res Public Health 2021; 18: 8751. doi: 10.3390/ijerph18168751.

[21] Nelsen RB. An Introduction to Copulas. Second Edition. USA:: Springer 2006.

[22] Masarotto G VC. Gaussian copula marginal regression. Electronic Journal of Statistics 2012; 6: 1517–49.

[23] Ferreira PH, Fiaccone RL, Lordelo JS, et al. Bivariate Copula-based Linear Mixed-effects Models: An Application to Longitudinal Child Growth Data. TEMA (São Carlos) 2019; 20: 37–59. doi: 10.5540/tema.2019.020.01.037.

[24] Cristiano Varin GM. Gaussian Copula Regression in R. Journal of Statistical Software 2017; 77: 1–26.

[25] Johnson JPJ, Arumugam R, Karuppusami R, et al. Intraoperative administration of systemic/epidural/intrathecal morphine on the quality of recovery following substitutional urethroplasty with buccal mucosal graft: A randomized control trial. Journal of Anaesthesiology Clinical Pharmacology 2022; 38: 537. doi: 10.4103/joacp.JOACP_589_20.

[26] Zeger SL, Liang KY. An overview of methods for the analysis of longitudinal data. Stat Med 1992; 11: 1825–39. doi: 10.1002/sim.4780111406.

[27] David J H, Martin J C. Practical longitudinal data analysis. 1st edn. London: Chapman and Hall: 1996.

[28] Hubbard AE, Ahern J, Fleischer NL, et al. To GEE or not to GEE: comparing population average and mixed models for estimating the associations between neighborhood risk factors and health. Epidemiology 2010; 21: 467–74. doi: 10.1097/EDE.0b013e3181caeb90.

[29] Twisk, J. Applied Longitudinal Data Analysis for Epidemiology: A Practical Guide. (2nd ed.). Cambridge:: Cambridge University Press 2013. doi: 10.1017/CBO9781139342834.

[30] Cnaan A, Laird NM, Slasor P. Using the general linear mixed model to analyse unbalanced repeated measures and longitudinal data. Stat Med 1997; 16: 2349–80. doi: 10.1002/(sici)1097-0258(19971030)16:20<2349::aid-sim667>3.0.co;2-e.

[31] Ballinger GA. Using Generalized Estimating Equations for Longitudinal Data Analysis. Organizational Research Methods 2004; 7: 127–50. doi: 10.1177/1094428104263672.

[32] Ghisletta P, Spini D. An Introduction to Generalized Estimating Equations and an Application to Assess Selectivity Effects in a Longitudinal Study on Very Old Individuals. Journal of Educational and Behavioral Statistics 2004; 29 (4): 421–37.

[33] Westgate PM, Burchett WW. A Comparison of Correlation Structure Selection Penalties for Generalized Estimating Equations. Am Stat 2017; 71: 344–53. doi: 10.1080/00031305.2016.1200490.

[34] Scott JM, deCamp A, Juraska M, et al. Finite-sample corrected generalized estimating equation of population average treatment effects in stepped wedge cluster randomized trials. Stat Methods Med Res 2017; 26: 583–97. doi: 10.1177/0962280214552092.

[35] Zeger SL, Liang KY, Albert PS. Models for longitudinal data: a generalized estimating equation approach. Biometrics 1988; 44: 1049–60.

[36] Sun J, Frees E, Rosenberg M. Heavy-tailed longitudinal data modeling using copulas. Mathematics and Economics 2008; 42: 817–30. doi: 10.1016/J.INSMATHECO.2007.09.009.

[37] Madsen L, Fang Y. Joint Regression Analysis for Discrete Longitudinal Data. Biometrics 2011; 67: 1171–5. doi: 10.1111/j.1541-0420.2010.01494.x.

[38] Escarela G, Rodríguez CE, Núñez-Antonio G. Copula modeling of receiver operating characteristic and predictiveness curves. Stat Med 2020; 39: 4252–66. doi: 10.1002/sim.8723.

[39] Jaman A, Latif MAHM, Bari W, et al. A determinant-based criterion for working correlation structure selection in generalized estimating equations. Stat Med 2016; 35: 1819–33. doi: 10.1002/sim.6821.