COMPARING LINEAR MIXED MODEL AND GENERALIZED ESTIMATING EQUATIONS TECHNIQUES FOR LONGITUDINAL DATA

Kamal Kishore*

Dr. D.K. Subbakrishna**

Dr. P. Marimuthu***

Abstract

The defining feature of longitudinal studies is correlated nature of data as response variable from same subjects are collected repeatedly on subsequent occasions. Moreover, a peculiar characteristic of longitudinal data is missing pattern. Mixed effect models (MEM) and generalized estimating equations (GEE) are two advanced statistical techniques which are popular among researchers for analysis of longitudinal data. These techniques were used to analyze longitudinal dataset on 95 subjects measured on six occasions. Although, these techniques are not comparable in general due to the different assumptions they make about the data. MEM and GEE techniques are comparable when fitted to same longitudinal data. The main focus of the study lies in the group growth trajectory. These approaches were compared on the basis of initial status and rate of growth of parameters, which are representative of a growth profile. Applied researchers are often confused as to which method performs better under different conditions so as to use them for the analysis of data. An attempt is made to demonstrate the differences and similarities in both the techniques for analyzing longitudinal data with and without missing pattern in the data.

Keywords: Longitudinal data, MEM, GEE, Traditional methods

* Assistant Professor, Institute of Management, Christ University, Bengaluru

** Professor, Department of Biostatistics, NIMHANS, Bengaluru

*** Additional Prof., Department of Biostatistics, NIMHANS, Bengaluru

Introduction

There has been tremendous growth in interest and statistical methodology for the longitudinal data analysis from last few years. One important reason for wide use of longitudinal studies can be attributed to its use to address the substantive research question about change over a period of time. Moreover, popularity is on the rise due to tremendous growth in literature and computational power at hands of end users. This growth leads to integration of advanced statistical techniques in popular software, thereby increasing the application of these methods.

Mixed effect models (MEM) and generalized estimating equations (GEE) approaches provide researchers with powerful and flexible analytic tools for the characteristics displayed by longitudinal data. MEM is an extension of the regression model in which dependent nature of subsequent observations from same subject are accomplished by introducing random effects. Moreover, random effects partitions total variability into within subject variability and between subject variability. MEM are very popular among the applied researchers and these are nicely described in literature by several authors (Fitzmaurice et al. 2003; Singer and Willett 2003; Twisk 2004; Fitzmaurice et al., 2009). GEE is a quasi-likelihood approach (Liang and Zeger, 1986; Zeger and Liang, 1986) which can pursue statistical models by making assumptions about the link function and the relationship between the first two moments, but without specifying the complete distribution of the response. GEE models are quite different than MEM, and these are described in literature by several authors (Diggle et al., 2002; Hardin and Hilbe, 2003; Liang and Zeger, 1993; Zeger et al., 1988). In most of applied research, the interest lies in the shape of population growth trajectory rather than individual growth trajectory over time and correlation among repeated measures. In this paper, an attempt is made to compare the performances of MEM and GEE, when research interest lie in group growth trajectory.

Before development of these advanced statistical techniques, most of the analysis for longitudinal data was carried out with traditional methods such as paired t-test and RM-ANOVA. An important characteristic of using advanced techniques are the way they handle dependence and missingness as compared to traditional methods in longitudinal data. When the subjects are repeatedly measured some of them miss schedules and some may permanently drop out from study due to one or the other reason. Unlike the traditional methods such as paired t-test, RM-ANOVA technique, these statistical techniques use all the subjects for analysis even if response is available for the single time point. Moreover, these techniques can be applied efficiently when outcome variable is categorical in nature.

MEM and GEE approach was developed in the beginning for different type of longitudinal outcome variable. MEM was initially developed for the analyses of normally continuous outcome longitudinal or clustered variable (Laird and Ware 1982; Bryk and Raudenbush 1992; Goldstein 1999) but were less familiar for analyses of non-normal correlated data. During same time, a quasi-likelihood estimation procedure known as generalized estimating equations (GEE), first introduced by Liang and Zeger(1986; Zeger and Liang 1986) have become very popular to estimate regression coefficients for analysis of longitudinal categorical data. These boundaries are now blurring due to growth in literature as both procedures are applicable for continuous as well as categorical data.

The idea of change and measurement of change is an intriguing concept which fascinate researchers from generations. This article compare performance of MEM and GEE techniques to measure change. Linear growth trajectory and quadratic growth trajectory are two commonly used assumptions to measure change. The linear growth trajectory assumes monotonic increase whereas for quadratic growth trajectory change increases upto a certain time and then levels off gradually.

Methodology

A MEM is appropriate statistical model for longitudinal data as it takes dependence into accounts through random effects. These models can be represented in mathematical form as:

$$Y_i = X_i\beta + Z_i b_i + \varepsilon_i, \quad i = 1,2, \dots \dots N \qquad (1)$$

where $Y_i = (y_{i1}, y_{i2}, \dots \dots, y_{in_i})'$, $X_i = (x_{i1}, x_{i2}, \dots \dots, x_{in_i})'$ is a n_ix p design matrix of known covariates, β is a p x 1 vector of fixed regression parameters, $Z_i$is a $n_i$xq design matrix for random effects, b_i is a q x 1 vector of random regression coefficients distributed as N(0, ∑) and b_is are mutually independent, and ε_i is n_i x 1 vector of random errors distributed as N(0, $\sigma^2 In_i$) and is independent of b_is. The variance-covariance matrix ∑ captures the degree of heterogeneity of subjects. It is important to note that $E(Y_{ij}) = E[E(Y_{ij}|b_i)] = x'_{ij}\beta$ and therefore marginal and conditional parameter are equal.

The MEM for binary outcome variable is an extension of MEM and generalized linear models for correlated non-Gaussian outcome variable. MEM for binary outcome variable can be specified with exponential family distribution for outcome variable, a link function and random effect structure. A MEM for binary outcome variable can be expressed with logit link

as follows:

$$E(Y_{ij}) = E[E(Y_{ij}|b_i)] = E\left[\frac{\exp(x'_{ij}\alpha + z'_{ij}b_i)}{1+\exp(x'_{ij}\alpha + z'_{ij}b_i)}\right] = \frac{\exp(x'_{ij}\alpha)}{1+\exp(x'_{ij}\alpha)}$$

GEE is a quasi-likelihood approach (Liang and Zeger, 1986; Zeger and Liang, 1986) which can pursue statistical models by making assumptions about the link function and the relationship between the first two moments, but without specifying the complete distribution of the response. A GEE model for longitudinal data has three part specification:

- Mean of each response is assumed to depend on the covariates through link function.

$$g(\mu)_{ij} = x'_{ij}\beta ,$$

- The conditional variance of y_{ij}, given the covariates is

$$Var(y_{ij}) = \emptyset\ var(\mu_{ij}),$$

  where Ø is known or estimated scale parameter and $var(\mu_{ij})$ is known variance function.

- With-in subject association among the vectors of repeated responses over time.

Data and Analysis

The data for the present study is utilized from Neurological performance of cohort consist of 95 HIV-1 infected individuals observed over $2^{1/}{}_2$ years in southern India. In the dataset the subjects were recruited at baseline and followed-up 5 times at six monthly intervals. The objective of research is to investigate the change in neuro-psychological performance over a period of time and the factors that influence change in group. The explanatory variables are either continuous or categorical and time dependent or time independent in nature.

The two advanced statistical techniques MEM and GEE for longitudinal outcome variable with continuous and categorical outcome variable were applied to neuro-psychological performance. The neuro-psychological performance was dichotomized in a way where each measurement of upper tertiles compared to two lower tertile. The dataset for comparing the two techniques were analyzed using “geepack”(Generalized & Equation, 2012; Halekoh, Højsgaard, & Yan, 2006; Højsgaard, 2011) and “lme4” (D Bates, Maechler, Bolker, & Walker, 2013; Douglas Bates, 2011) package of free and flexible R-software..

MEM was developed initially for continuous data, whereas GEE was primarily developed for categorical data. The developments in literature lead to extension of both these techniques for analysis of categorical and continuous data. MEM and GEE generate subject-specific parameter and population specific parameter estimate respectively. Population specific parameter estimates can be obtained from subject specific parameter estimates however, population specific estimates of MEM for categorical data are biased (Agresti 2002).

One of the major challenges in dealing with longitudinal studies is of missing data and the major difference between MEM and GEE is the way they handle missing data. In this paper original dataset analysis is followed by selecting subsets (incomplete datasets) of data from original data. The incomplete datasets were obtained from complete datasets by deleting approximately 21% (N=40) of the observations from fourth occasion. MCAR is the strongest assumption for the data which is rarely met by data.Data following MCAR mechanism can be thought of random sample of the complete data. MCAR data was generated by omitting values completely at random from complete data. Data are said to be missing at random (MAR) when the probability that response are missing depends on the set of observed responses, but is unrelated to the specific missing value that in principle, should have been obtained (G. M. Fitzmaurice, Laird, & Ware, 2004). MAR mechanism occurs more frequently in longitudinal studies. One of the major consequences of MAR is that complete cases are not random samples and it can lead to biased estimates of change in mean response over time. MAR data was generated by arranging the value in ascending order at 3rd occasion and then subsequently removing the lowest 40 values from the fourth occasion onwards. The data is categorized before omitting the values and then same values are omitted which were omitted for continuous outcome variable.

Results

There were total 95 subjects at the beginning of study. Demographic profile of the group suggests comparable gender distribution of male (n=58; 61%) and female (n=37; 39%). The subjects were in the age group of 20-45 years, 58% of the subjects were married and most of them were from urban population (n=40; 42%) followed by rural (n=26; 27%). Further, study group consisted of majority of subjects from high literacy group (n=60; 63%) and most of them were from nuclear families (n=52; 55%). Longitudinal data used for analysis showed significant correlation over a period of time as can be seen below in the table (1).

Table 1. Observed inter-occasion correlation coefficient for outcome variable

	Time1	Time2	Time3	Time4	Time5	Time6
Time1	-	0.807**	0.784**	0.582**	0.467**	0.6**
Time2		-	0.792**	0.697**	0.609**	0.645**
Time3			-	0.731**	0.599**	0.643**
Time4				-	0.685**	0.565**
Time5					-	0.585**
Time6						-

Individual and Group Profile

The facts about mean response trajectory can be visualized from figure (1). It may be noted that mean neuropsychological performance increases smoothly upto certain period and then decreases. This figure indicates that response is not linear with time and polynomial growth curve can be a possibilty to account for curviliear nature of change. However, the final decision to include polynomial trajectory is to be made on the basis of testing procedure. In order to visualize how different persons change over a period of time empirical growth plots of 12 randomly selected subjects are shown here in figure 2.

Figure (1)

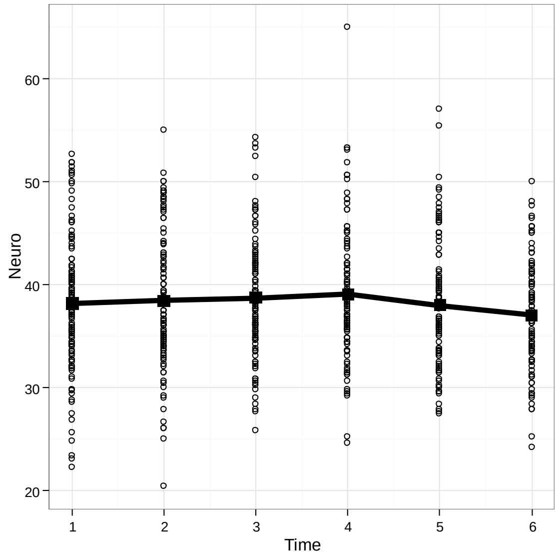
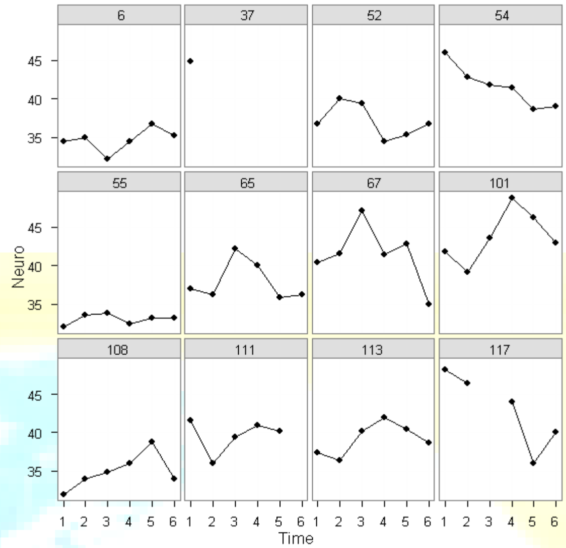


Figure (2)



Comparing Performances of MEM and GEE for continuous outcome variable

The main goal of the study was to infer, how the covariates are related to change in outcome variable. There were not any interest in interactions of covariates with time.The covariates of interest to study are gender, education level, income, locality, family, BDI score, Marital status, MMSE score and CD4 count at baseline. Mathematically a model for continuous outcome for linear growth trajectory can be wriiten as:

$$Y_{ij} = \beta_0 + \beta_1(Time) + \beta_3(Gender) + \beta_4(Edu) + \beta_5(Income) + \beta_6(Locality) + \beta_7(Family) + \beta_8(BDI) + \beta_9(MMSE) + \beta_{10}(CD4C) + \beta_{11}(Mstatus) + \varepsilon_{ij} \qquad (1)$$

A MEM with random intercept is equivalent to exchangeable correlation structure in GEE and MEM with random intercept and random slope is equivalent to unstructured correlation in GEE. The precision of parameter estimates are inversely proportional to standard error, more the standard error lesser the precision and vice–versa. The results obtained after application of MEM and GEE for linear growth trajectory and quadratic growth trajectory for continuous and categorical outcome variable are displayed and discussed subsequently. Missing data is almost always part of longitudinal studies, thus two incomplete data (MCAR and MAR) are generated so as to observe how these techniques handle missing data.

Table (2)

Model		Random Intercept			Random Intercept and Slope		
		Complete	MCAR	MAR	Complete	MCAR	MAR
Intercept	MEM	33.93 (0.87)	33.66 (0.88)	33.53 (0.95)	34.15 (0.92)	33.92 (0.93)	33.74 (1.00)
	GEE	33.93 (0.82)	33.67 (0.82)	33.52 (0.90)	33.51 (0.84)	33.55 (0.78)	33.65 (0.88)
Slope	MEM	-0.15 (0.09)	-0.13 (0.10)	-0.15 (0.11)	-0.15 (0.11)	-0.12 (0.12)	-0.14 (0.13)
	GEE	-0.14 (0.10)	-0.13 (0.12)	-0.15 (0.13)	-0.06 (0.10)	-0.11 (0.12)	-0.20 (0.13)
Gender	MEM	5.08 (0.88)	5.07 (0.87)	5.27 (0.91)	4.93 (0.87)	4.83 (0.86)	5.05 (0.90)
	GEE	5.08 (0.90)	5.07 (0.89)	5.28 (0.92)	4.59 (0.87)	4.64 (0.86)	4.38 (0.89)
Education	MEM	4.25 (0.89)	4.33 (0.88)	4.69 (0.94)	3.99 (0.88)	4.11 (0.87)	4.50 (0.92)
	GEE	4.25 (0.87)	4.33 (0.89)	4.68 (0.90)	4.18 (0.85)	4.70 (0.80)	4.72 (0.87)

In the table (2), parameter estimates and their respective standard errors obtained by MEM and GEE techniques are presented. The exchangeable and unstructured correlation structure models of GEE with equivalent representative models from MEM were compared against each other. The parameter estimates and their standard errors are almost same with GEE and MEM techniques. In the literature it is emphasized that a major difference between MEM and GEE is the way they handle missing data, the former can handle data with MAR mechanism but not the later one. From table (2) it can be seen that GEE estimates and their standard errors are nearby to MEM for data with MAR mechanism. There are some differences in parameter estimates and their standard errors but, these differences can be attributed to maximum likelihood estimation method of MEM and quasi-likelihood estimation method for GEE. The GEE model is more flexible than MEM in making assumptions about the data.

Table (3) is displaying the results for quadratic growth curve models. The model can be obtained by assuming quadratic growth trajectory and adding the same in equation (1). It can be observed from the table that both the techniques are giving almost the same results. Moreover, standard error is more for data with MAR mechanism.

Table (3)

Model		Random Intercept			Random Intercept and Slope		
		Complete	MCAR	MAR	Complete	MCAR	MAR
Intercept	MEM	31.05 (1.03)	30.65 (1.06)	29.99 (1.15)	31.27 (1.05)	30.97 (1.09)	30.14 (1.18)
	GEE	31.05 (1.00)	30.65 (1.00)	29.98 (1.10)	31.43 (0.98)	30.87 (0.95)	30.57 (1.14)
Slope	MEM	2.02 (0.42)	2.16 (0.47)	2.59 (0.52)	2.02 (0.40)	2.10 (0.45)	2.59 (0.50)
	GEE	2.02 (0.43)	2.16 (0.44)	2.59 (0.51)	1.87 (0.39)	1.98 (0.42)	2.30 (0.56)
Quadratic	MEM	-0.31 (0.06)	-0.33 (0.07)	-0.39 (0.07)	-0.31 (0.05)	-0.32 (0.06)	-0.39 (0.07)
	GEE	-0.31 (0.06)	-0.33 (0.06)	-0.39 (0.07)	-0.29 (0.05)	-0.29 (0.06)	-0.34 (0.08)
Gender	MEM	5.08 (0.88)	5.05 (0.87)	5.25 (0.92)	4.93 (0.87)	4.82 (0.86)	5.07 (0.90)
	GEE	5.08 (0.90)	5.05 (0.89)	5.26 (0.92)	4.57 (0.87)	4.58 (0.86)	4.43 (0.89)
Education	MEM	4.25 (0.89)	4.32 (0.88)	4.56 (0.94)	3.99 (0.88)	4.11 (0.87)	4.40 (0.92)
	GEE	4.25 (0.87)	4.32 (0.85)	4.55 (0.90)	4.20 (0.85)	4.64 (0.80)	4.33 (0.86)

The positive parameter estimate for slope component and negative parameter estimate for quadratic component is indicator of growth rate in beginning and then decrease subsequently.It is emphasized in literature that GEE technique is robust to misclassification of correlation structure, but here this trend is also observed for MEM technique. From the table (3) it can be said that GEE and MEM estimates and their standard errors are very stable and almost near for Continuous outcome variable for both complete and incomplete datasets. Thus it was observed that GEE and MEM estimates and their standard errors are not very different from each other for complete and Incomplete (MCAR and MAR) datasets, when the growth trajectories assumed were linear and quadratic.

Comparing Performances of MEM and GEE for binary outcome variable

The model building is done to summarize the important characteristics of data with parsimonious model which should have all the relevant parameters. Up to this point, the repeated measures analysis of the variables was restricted for continuous response variable. However, in practice, categorical variables are also very common in applied research. Because the response variable can take one of two values it cannot be interpreted like continuous response variable. The model for binary response variable was adapted to predict the probability of positive response (D. M. Bates, 2010). A data set was analyzed in which subjects are assesed over a period of time for binary outcome. A marginal logistic regression model for binary response variable with all the covariates of interest is represented by equation 2. The interaction effect among covariates was not of interest and was not assumed for estimation process.

$$logit[p(Y_{ij}=1)]=\beta_0+\beta_1(Time_{ij})+\beta_3(Gender)+\beta_4(Edu)+\beta_5(Income)+\beta_6(Locality)+$$
$$\beta_7(Family)+\beta_8(BDI)+\beta_9(MMSE)+\beta_{10}(CD4C)+\beta_{11}(Mstatus)+\varepsilon_{ij} \quad (2)$$

The parameter estimates for categorical data for linear growth trajectory are displayed in table (4). It can be observed from table that unlike estimates and their standard errors for continuous data, standard error estimates obtained with GEE technique is always better. GEE estimates for data with MAR mechanism holds for categorical data also. The criticism in literature against usage of GEE for data missing at random is not supported by this study. According to this study parameter estimates and their standard errors obtained by GEE are as efficient as estimates obtained by MEM.

Table (4)

Model		Random Intercept			Random Intercept and Slope		
		Complete	MCAR	MAR	Complete	MCAR	MAR
Intercept	MEM	-0.80 (0.51)	-0.99 (0.52)	-1.04 (0.61)	-0.41 (0.58)	-0.69 (0.60)	-0.79 (0.72)
	GEE	-0.59 (0.34)	-0.69 (0.35)	-0.67 (0.39)	-0.50 (0.33)	-0.66 (0.35)	-0.76 (0.40)
Slope	MEM	-0.06 (0.07)	-0.04 (0.08)	-0.04 (0.09)	-0.12 (0.08)	-0.09 (0.09)	-0.11 (0.11)
	GEE	-0.04 (0.04)	-0.03 (0.05)	-0.03 (0.06)	-0.03 (0.04)	-0.03 (0.05)	-0.01 (0.06)
Gender	MEM	2.31 (0.52)	2.25 (0.53)	2.49 (0.59)	2.28 (0.53)	2.20 (0.54)	2.67 (0.65)
	GEE	1.51 (0.35)	1.47 (0.35)	1.54 (0.36)	1.48 (0.34)	1.45 (0.34)	1.55 (0.37)
Education	MEM	2.14	2.22	2.43	2.02	2.19	2.61

		(0.05)	(0.51)	(0.58)	(0.51)	(0.52)	(0.63)
	GEE	1.45 (0.32)	1.48 (0.33)	1.51 (0.35)	1.33 (0.32)	1.49 (0.33)	1.52 (0.36)

Similarly parameter estimates and standard errors obtained with both the techniques for quadratic growth trajectory are displayed in table (5).

Table(5)

Model		Random Intercept			Random Intercept and Slope		
		Complete	MCAR	MAR	Complete	MCAR	MAR
Intercept	MEM	-2.72 (0.71)	-2.52 (0.74)	-2.99 (0.87)	-2.49 (0.77)	-2.32 (0.82)	-2.88 (1.02)
	GEE	-1.67 (0.43)	-1.58 (0.46)	-1.64 (0.45)	-1.45 (0.37)	-1.41 (0.43)	-1.68 (0.45)
Slope	MEM	1.35 (0.35)	1.09 (0.37)	1.40 (0.45)	1.34 (0.36)	1.08 (0.39)	1.43 (0.49)
	GEE	0.82 (0.22)	0.67 (0.24)	0.78 (0.24)	0.71 (0.19)	0.56 (0.22)	0.81 (0.24)
Quadratic	MEM	-0.20 (0.05)	-0.16 (0.05)	-0.21 (0.06)	-0.20 (0.05)	-0.17 (0.05)	-0.22 (0.07)
	GEE	-0.12 (0.03)	-0.10 (0.03)	-0.12 (0.04)	-0.11 (0.03)	-0.08 (0.03)	-0.12 (0.03)
Gender	MEM	2.43 (0.55)	2.33 (0.55)	2.64 (0.63)	2.45 (0.57)	2.27 (0.56)	2.80 (0.68)
	GEE	1.44 (0.35)	1.42 (0.35)	1.44 (0.36)	1.43 (0.34)	1.39 (0.34)	1.44 (0.36)
Education	MEM	2.25 (0.53)	2.29 (0.53)	2.52 (0.62)	2.17 (0.54)	2.25 (0.54)	2.68 (0.67)
	GEE	1.38 (0.32)	1.43 (0.33)	1.38 (0.35)	1.26 (0.32)	1.45 (0.32)	1.40 (0.35)

The parameter estimates and their standard errors obtained with GEE technique for categorical outcome are always lower than MEM, whereas this trend was not noticed for continuous outcome variables. The standard errors are lower and more stabilized for parameter estimates obtained with GEE technique. The incomplete data had marginal effect on the parameter estimates and their standard errors. The standard errors are higher for data with MAR mechanism for both GEE and MEM. It is surprising to note that, parameter estimates and their standard errors with MEM and GEE techniques are quite different for binary outcome variables compared to continuous outcome variable. More surprising are the results for GEE technique applicable to data with MAR mechanism as these are better than results obtained with MEM. Thus, it was observed that parameter estimates and their standard errors obtained by GEE technique are more stable and efficient than MEM technique for categorical data. Moreover, surprisingly same trend was observed for incomplete data in general and data with MAR mechanism in particular.

Discussion

Both MEM and GEE techniques are highly suitable and preferred among researchers for analyzing longitudinal data. The question arises: which one of these techniques is better? In this paper these two popular statistical techniques for analysis of longitudinal data are compared and discussed. MEM and GEE techniques are very appealing as they include time and their transformations to infer about group growth trajectories. Moreover, both techniques use all the available data for analysis even if data is not available for all but one occasion. The MEM is an extension of linear regression which was extended by adding random effect in regression for longitudinal continuous outcome. Whereas during same time, GEE technique was developed for categorical data where generalized linear models (GLM) was extended by incorporating correlation structure. MEM and GEE approaches were used on empirical longitudinal dataset for comparison of two approaches.

The GEE approach does not make distributional assumptions because estimation of population-averaged model depends on correct specification of few aspects of observed data and it does not depend on the entire data generating distribution. GEE models for mean response depend on the predictors of interest rather than on random effects or previous responses. It requires only a regression model for mean response and does not require distributional assumptions for the data (G. M. Fitzmaurice et al., 2004). On the other hand MEM requires correct specification of distributional assumptions about random effects besides usual data distributional assumptions. The parameter estimates for fixed effect can be misleading and may lead to biased inference as they depends on correct specification of non-identifiable latent random effect for which even large sample size do not help (Hubbard et al., 2010)

Since, interpretation of parameter estimates is different for population-averaged model as compared to subject specific models, their selection is of vital importance to address question of scientific interest. The MEM models are subject-specific whereas GEE are population-averaged model. The researcher should clearly state about the interest of analysis. In case of continuous outcome variable subject-specific estimates can be averaged to get population averaged estimates but same cannot be done for categorical data without introducing bias. Both statistical techniques resulted in comparable parameter estimates and their standard errors for continuous outcome variable. GEE estimates are more efficient for categorical outcome variable but these are population average estimates. Whenever, the research interest lies in individual change MEM are preferred choice.

Missing data is one of the major challenges in longitudinal studies. Both techniques use all the available data for analysis. In literature, it is emphasized that GEE can handle data with MCAR mechanism, whereas MEM can handle data with MAR mechanism. It is observed here that parameter estimates and standard errors with GEE and MEM are slightly different for complete and incomplete continuous datasets. Further, it was observed that there was not much difference in both the techniques for random missing data (MCAR) as compared to dataset with selective missingness (MAR). The same trend was noticed for binary outcome variable for both the techniques but estimates obtained with GEE technique were more efficient as compared to MEM estimates. These results are rather surprising in contrast to existing literature as GEE performed consistently for selective missingness (MAR) in case of both continuous and categorical outcome variable.

Conclusions

GEE and MEM techniques for continuous outcome variable leads to similar results for complete and incomplete datasets. Any technique can be applied when interest lies in population averaged trajectory. MEM is preferred when research interest is in individual growth trajectory. For categorical growth outcome variable GEE technique is more efficient as compared to MEM technique. Moreover, MEM approach brings bias when averaged to group growth trajectory. It is to be used only when research interest is in individual growth trajectory.

Bibliography

Bates, D. M., 2010, lme4: Mixed-effects modeling with R. URL http://lme4. R-Forge. R-Project. Org/book.

Bates, D., 2011, Mixed models in R using the lme4 package Part 1 : Linear mixed models with simple , scalar random effects. Development, pp 1–27.

Bates, D., Maechler, M., Bolker, B., & Walker, S., 2013. lme4: Linear mixed-effects models using Eigen and S4. R Package Version, pp 0–1.

Bryk, A. S., & Raudenbush, S. W.,(1992),Hierarchical Linear Models. 1st. Ed. Newbury Park, CA: Sage Publications.

Diggle, P., Heagerty, P. J., Liang, K., & Zeger, S. L., (2002). Analysis of longitudinal data, 3rd. Ed., Oxford University Press, USA.

Fitzmaurice, G., Laird, N., & Ware, J., (2003). Applied longitudinal analysis, 2nd. Ed., John Wiley & Sons.

Goldstein, H.,(1999). Multilevel Statistical Models, 3rd. Ed., London: Edward Arnold.

Halekoh, U., Højsgaard, S., & Yan, J., 2006, The R package geepack for generalized estimating equations. Journal of Statistical Software, 15(2), pp 1–11.

Hardin, J. W., & Hilbe, J. M.,(2003). Generalized Estimating Equations, 2nd. Ed., Chapman and Hall/CRC.

Højsgaard, S. (2011). On the usage of the geepack, (1), pp 1–4.

Hubbard, A. E., Ahern, J., Fleischer, N. L., Van der Laan, M., Lippman, S. A., Jewell, N., Satariano, W. A., 2010, To GEE or not to GEE: comparing population average and mixed models for estimating the associations between neighborhood risk factors and health. Epidemiology, 21(4), pp 467–474.

Laird, N. M., & Ware, J. H., 1982, Random-effects models for longitudinal data. Biometrics, pp 963–974.

Liang, K., & Zeger, S. L., 1986, Longitudinal Data Aanalysis Using Generalized Linear Models. Biometrika, 73(1), pp 13–22.

Liang, K.-Y., & Zeger, S. L., 1993, Regression analysis for correlated data. Annual Review of Public Health, 14(1), pp 43–68.

Singer, J. D., & Willett, J. B., (2003). Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence, 1st. Ed., Oxford University Press.

Twisk, J. W. R., 2004, Longitudinal data analysis: A comparison between generalized estimating equations and random coefficient analysis. European Journal of Epidemiology, 19(8), pp 769–76.

Zeger, S. L., & Liang, K.-Y., 1986, Longitudinal data analysis for discrete and continuous outcomes,Biometrics, pp 121–130.

Zeger, S. L., Liang, K.-Y., & Albert, P. S., 1988, Models for longitudinal data: a generalized estimating equation approach. Biometrics, pp 1049–1060.