ISSN: 2322 - 0902 (P)
ISSN: 2322 - 0910 (O)

International Journal of Ayurveda and Pharma Research

Research Article

GENERALIZED ESTIMATING EQUATION IN PHARMACOLOGICAL STUDY

A. Tripathi[1]*, M. Bora[2], Lalrinpuia[2], K. Mukherjee[3], S. N. Upadhaya[4]

*[1]Statistical Officer, [2]Research Officer, [3]Senior Research Fellow, [4]Assistant Director, Central Ayurveda Research Institute Drug Development, Bidhannagar, Kolkata, West Bengal, India.

ABSTRACT

Generalized Estimating Equation (GEE) deals with such data that do not have normal distribution in case of repeated experiment and has better properties as compared to rANOVA. With the objective to describe the use of GEE in pharmacological study this endeavour started. GEE described hereby using rat data. Four correlation structured were taken in GEE. It was found Independent/Exchangeable structure best suited with data. The model fit on data was assessed graphically as well. The trend line of repeated data for all cases (rats) were fallen in 95 % bound of predicted model. The model gave the average weight of rat 226.51 gram with start of experiment and it increases 8.67gram per week after feeding high fat diet.

KEYWORDS: Generalized Estimating Equation (GEE), Pharmacological study,

INTRODUCTION

Generalized Estimating Equation (GEE) is a method of longitudinal data analysis [1]. GEE is a population level approach to estimate the parameters of model and provides semi parametric way to longitudinal data analysis. GEE is used by researcher frequently in place of repeated measure analysis of variance (rANOVA) in case of time followed data [2]. rANOVA compares the mean of outcome variable at different time point or between group. E.g. In case of Randomized Controlled Trial (RCT) time being efficacy of both drug (New drug and standard drug) and difference between drug efficacy are compared . rANOVA is well established method of longitudinal data analysis in case of clinical trial but it has many assumption about data where it can be applicable like the outcome variable should be quantitative variable with normal distribution and covariates are categorical variables[3]. It requires that the outcome have constant variance across time points and constant correlation between any two time points. rANOVA does not perform well when sample size is small, data is not normal and loss to follow up of cases are high[4]. GEE have better features as compare to rANOVA in many circumstances [5]. GEE provide an overall measure of trend in longitudinal data analysis. It does not have many constraints in longitudinal data analysis as compared to rANOVA. The objective of GEE to provide the average change in response over time as well as impact of covariates on these changes[6]. GEE models the mean response as linear function. According to distribution of outcome variables many link function is there to link mean response as a linear function of covariates. In nutshell GEE is repeated measure analogue of Generalized Linear Model (GLM). GEE is preferred by researcher in case of toxicity/Pharmacological study where sample size is very low and distribution of data is not clear. In toxicity/Pharmacological study, Rats/ Mice /Guinea pig/ Rabbits etc. are being sacrificed, so due to ethical reasons very less no. of subject is enrolled by researcher. [7]

So the main objective of this study is to describe GEE such a way that a person who does not belong to statistical science can be easily apply this method in their research. Here pharmacological data is used by authors, which is generated in a pharmacological study. This Pharmacological study is concerned about change in body weight of mice after taking high fat diet.

Methodology

Suppose a pharamacological study has taken k (i= 1,2,…..,k) rats in the study and follow these rats to m_i time point. Y_{ij} denote the value of outcome variable for ith rat at jth time point. X_{ij} denote a t×1 vector of covariates/independent variables.Y_i =$(Y_{i1}, Y_{i2}$,….., $Y_{in_i})'$ is a outcome vector for the ith subject. μ_i=$(\mu_{i1}, \mu_{i2}$,….., $\mu_{in_i})'$ is a mean vector for ith subject where μ_{ij} is the corresponding jth mean. The responses are correlated within subject but independent in between subject. If f is a link function which is establish the relationship between covariate and response variable, β is the regression matrix. The

considered model builds a relationship in following way.

$$f(\mu_{ij}) = X'_{ij}\beta$$

The above model links the response variable to independent variable and whatever the distribution is followed by response variable link to independent variable by using this function. e.g. like in a study if researchers are interested to find out the change in proportion of people affected by disease at each follow up where certain drug intervene to cope up with disease so this is something the data where response variable measure in proportion and follow a binomial distribution. In this cases the link function take a form of logit function[8]. Following table give the examples of link functions:

Table No.-1: Link functions in GEE with examples

Distribution of Response/Outcome variable	Link Function	Example
Binomial	Logit, Probit	Proportion of persons getting free from eye infection at every week when treated by eye drop
Poisson	Log	Increment in Blood donation rate on successive blood camps
Normal	Identity	Body weight of rats on successive days after giving High fat diet
Gamma	Log	Health Cost Data in certain disease

The variance of response variable is calculated in next step in the form of variance-covariance matrix. Variance for each time point is taken as σ_j^2 and covariance is defined as $\rho\sigma_i\sigma_j$. Here ρ is the within subject correlation.

This model link the mean response to covariates like GLM but the special thing which is considered by this model is within subject correlation. Structure of correlation in model is taken by correlation matrix. Correlation matrix represents the within-subject dependencies. Its size is determined by the number of measurements. The important correlation structure are describing as follows:

1. Independent Structure: Repeated measurements are uncorrelated. This is quite tough to believe.
2. Autoregressive Structure: Repeated measurements have a first-order autoregressive relationship. The correlation between any two elements is equal to ρ for adjacent elements, ρ^2 for elements that are separated by a third, and so on. It means with consecutive repeated measurement the correlation become poor with the difference in repetition.
3. Compound symmetry structure: This structure has homogenous correlations between elements. It is also known as an exchangeable structure. This assumption strictly followed in the rANOVA.
4. Unstructured: This is a completely general correlation matrix.

The parameters of model are estimated by quasi-likelihood procedure[9] and the variance covariance matrix can be estimated by robust "sandwich" variance estimator. In this technique the variances of the parameters of interest are obtained robustly, and remain valid even if the working correlation matrix is not an estimate of the true correlation matrix[10]. However if correlation structure of responses is mis-specified the standard errors of parameters are not good. The value of quasi likelihood criterion (QIC) is considered as a measure for choosing between two correlation structures, given a set of model parameters [11]. Smaller value indicates the better model.

Result

Descriptive profile of selected rats shows in figure-1. The box plot of figure-1 shows the mean± 2SD of rats at each week. It is clear with the box plot with successive week the mean of body weight of rats as well as their variation increasing. It is also clear that this data is not suitable for rANOVA because rANOVA assume the constant variance at each follow up point.

GEE model applied on data with consideration of four correlation structure pattern (Table-2). Intercept as well as time/week (taken as covariate) has taken in the model. The intercept as well as effect/regression coefficient of time are nearly same for all models (for different correlation structure). The Independent structure and exchangeable structure models having the same parameter estimates. They also show the lowest QIC. Figure-2 shows the weight change trend for each rat at twelve different weeks, each blue line represent a rat. Independent/Exchangeable structure model taken as final and by using the parameter of this model predicted line are drawn. The main black line shows the predicted line. The two red lines show the upper and lower bound of the model. It is clear from

figure-2 that trend line of each rat come under the LB and UB of the predicted model. On the basis of parameter estimates of model the average weight of rat population at baseline is 226.51 gram and each successive week rat population gain average 8.67 gram.

Figure 1: Descriptive Profile of Selected rats in study

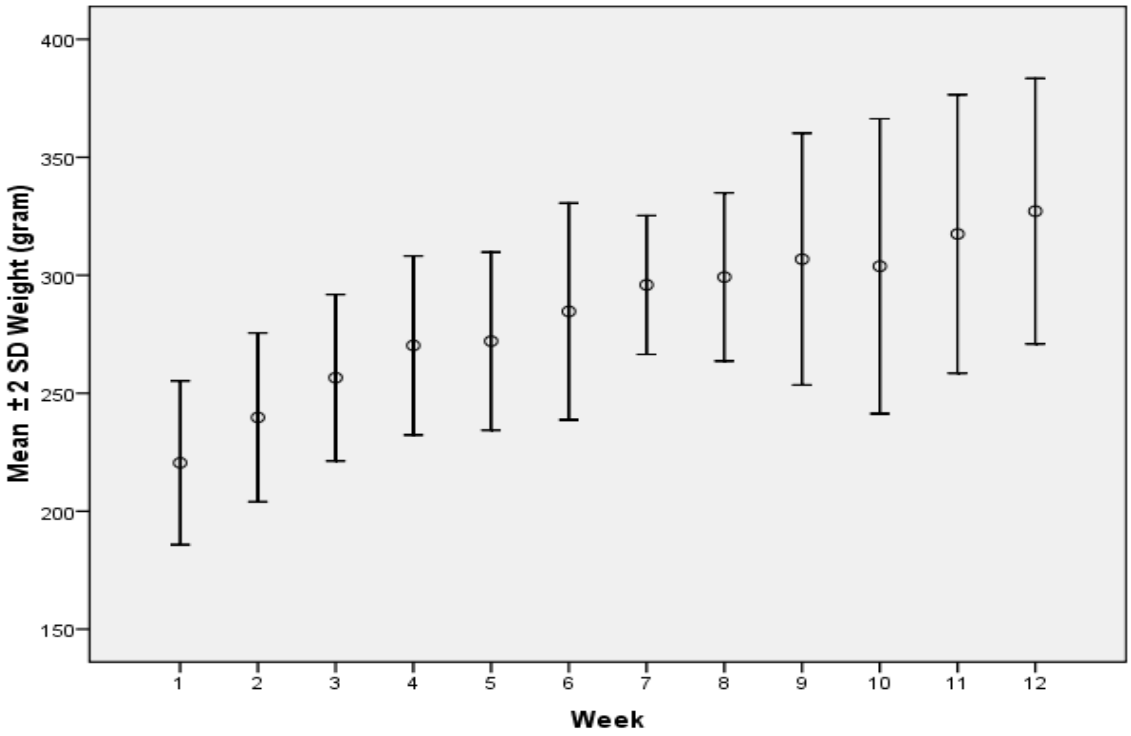
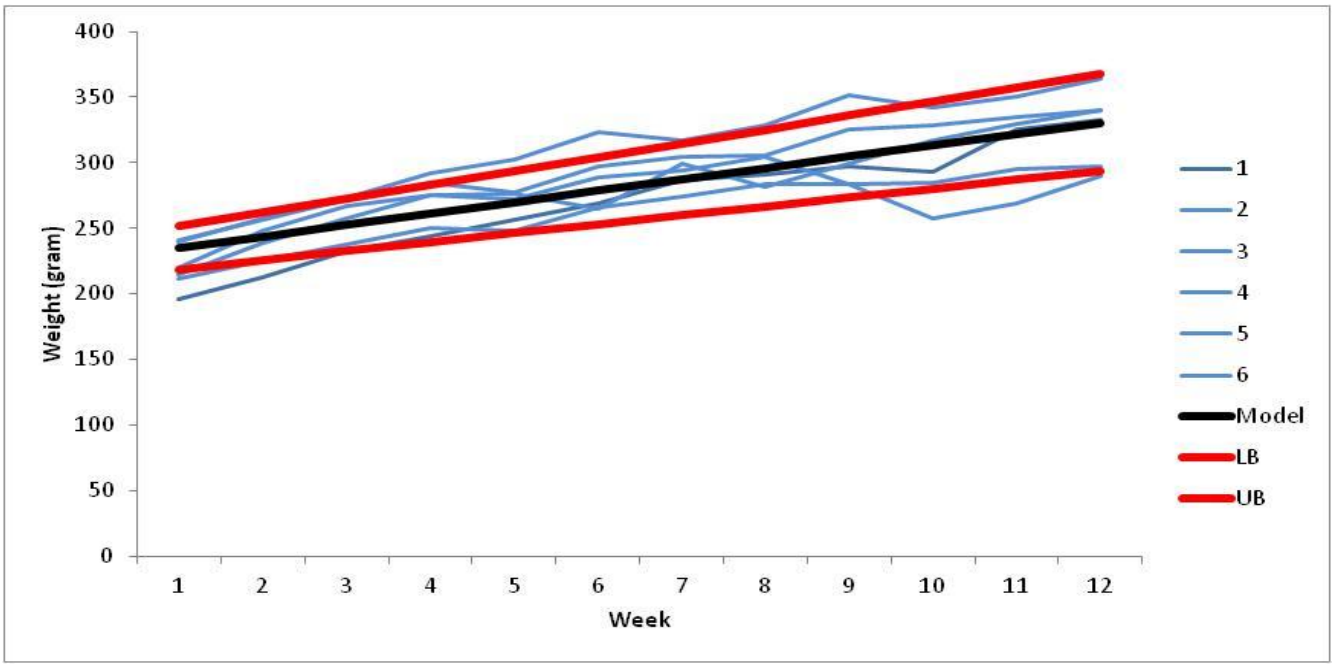


Table No.2: GEE model parameters with different correlation structure.

Correlation Structure	Intercept(I)	SE(I)	Effect of Time(T)	SE(T)	QIC
Independent	226.51	7.36	8.67	0.96	42380
Autoregressive	215.70	6.75	9.58	0.83	44870
Exchangeable	226.51	7.36	8.67	0.96	42380
Unstructured	211.66	8.68	9.67	0.77	48620

Figure 2: Prediction capacity of model



DISCUSSION

The present study concerned to describe the various aspect of GEE. A data of pharmacological study on rats had taken to describe the GEE procedure. The various correlation structures are taken in GEE model. The Independent/Exchangeable correlation structure model best suited to the data. There is a significant change found in trend of body weight increment after intervention of high fat diet in rats. GEE does not want the normality condition for response variable or covariates. GEE procedure capable to treat the missing data in follow-up study. It considers the missing data as a random missing and the missing data are just a random subset of the data. Though GEE has many good features but a strong model selection criterion is needed.

REFERENCES

1. Hardin J W and Hilbe J M, Generalized Estimating Equations, Chapman and Hall/CRC Press, Boca Raton, Fla, USA, 2003.
2. Armitage P, Berry G & Matthews J N S, Statistical Methods In Medical Research, Blackwell Science Publishing, 2002.
3. Krystal G (2004). "Move Over ANOVA". Arch Gen Psychiatry.61:310.doi:10.1001/archpsyc.61.3.310.
4. Oberfield D & Franke T, Evaluating the robustness of repeated measures analyses: The case of small sample sizes and nonnormal data. Behav Res (2013) 45 : 792-812.
5. Kreuger; T (2004). "A comparison of the general linear mixed model and repeated measures ANOVA using a dataset with multiple missing data points". Biological Research for Nursing. 6: 151–157.doi:10.1177/1099800404267682
6. Kung-Yee Liang and Scott Zeger (1986). "Longitudinal data analysis using generalized linear models". Biometrika. 73 (1): 13–22
7. Ghosh M N Vedasiromani J R, 2015. Fundamentals of experimental Pharmacology. Hilton & Company 109, College Street, Kolkata-700012.
8. Agresti A. (2002), Categorical Data Analysis, John Wiley &Sons.
9. Wedderburn R. W, "Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method," Biometrika, vol.61, pp. 439–447, 1974.
10. G. Kauermann and R. J. Carroll, "A note on the efficiency of sandwich covariance matrix estimation," Journal of the American Statistical Association, vol. 96, no. 456, pp. 1387–1396, 2001.
11. Pan, W. (2001), "Akaike's information criterion in generalized estimating equations," Biometrics, 57, 120-125.

Cite this article as:
A. Tripathi, M. Bora, Lalrinpuia, K. Mukherjee, S. N. Upadhaya. Generalized Estimating Equation in Pharmacological Study. International Journal of Ayurveda and Pharma Research. 2018;6 (5) :89-92.
Source of support: Nil, Conflict of interest: None Declared

***Address for correspondence**
Dr A Tripathi
Statistical Officer,
Central Ayurveda Research Institute
Drug Development, Bidhannagar,
Kolkata, West Bengal, India.
Email: ashoka07bhu@yahoo.in