A Bayesian Joint Analysis and Imputation Model for Longitudinal Data: An Application in Type 2 Diabetes Drug Effect Comparison

Atanu Bhattacharjee *

Ph.D (Statistics), M.Sc (Health Statistics), BSc Hons (Statistics), Maulana Azad Institute of Dental Science, New Delhi-110002, India

**** Corresponding Author;*** Email: atanustat@gmail.com

Abstract

Background: The level of serum creatinine is important affected parameter in presence of type 2 diabetes. The choice of type 2 diabetes drug therapy is crucial to control the serum creatinine level. The drug treatment effect can only be captured through repeated observations in the patients.

Objective: The aim of this work is to compare the drug treatment effect (i.e. "Metformin plus Pioglitazone" and "Gliclazide plus Pioglitazone") in presences of repeatedly measured missing observations to control serum cretinine levels in type 2 diabetes patients.

Method: The joint longitudinal modeling approach is applied to deal with missing observations. The presences of missing observations are assigned with missing at random and not random. The Markov chain Monte Carlo (MCMC) is used to carry out the iteration procedure.

Results: The "Metformin plus Pioglitazone" is found more effective to control serum creatinine in comparison to "Gliclazide with Pioglitazone". The joint longitudinal model with consideration of missing assumption proffers enhanced tool for inference on clinical trial data analysis.

Conclusion: The presence of missing observation is natural in repeated measurement. The tendency is to overlook the trial having observation and conclusion with missing observation. The elaborated method can be applied in other clinical trial problem to reduce the inconsistency due presence of missing observations.

Key words: serum creatinine, type 2 diabetes, Markov chain Monte Carlo

Introduction

The higher level of serum creatinine can affect kidney function. The skeletal muscle is insulin resistance and accumulate higher amount of glucose disposal after glucose infusion.[1] The level of serum creatinine is highly correlated with renal function[2] and type 2 diabetes.[3] The uncontrolled serum creatinine can lead the advance prostate cancer.[4,5] However, the contradictory results are also documented.[6,7,8] Therefore, the levels of serum creatinine measures are frequently considered in the clinical trials. The inferences about drug treatment effect can be fettered due to presence of drop-out of observations of the corresponding follow-up visits. This work is contributed to overcome the cumbersome occurred due to presences of missing observation of serum creatinine in repeated measurements.

In longitudinal study design, the control and experiment groups are measured repeatedly over the study period. The occurrence of missing observations is usual at all time points. Briefly, two types of reason can be figured out for the presence of drop-out of the repeated measurement. Firstly, the patients have not participated in the study after a certain times. Secondly, the drop-out occurs due to not filling up the questionnaire. The reason for not filling the questionnaire can be fault of both side i.e. Interviewer and Interviewee.

The different types of drop-outs are present in the longitudinal data analysis.[9] The type of drop-out can be classified through (i) MCAR (II) MAR and (III) MNAR.[10,11,12,13]

(I) If no reason can be found for the presence of drop-out, then it is known as MCAR.

(II) It is known as MAR, drop-out becomes dependent with previous observation but independent with current and future measurements.

(III) It is known as MNAR, if it be dependent on the current and future observations. It is also known as Informative. It is general tendency to analyze the trial data through the assumption of MAR. In MAR, the patients drop-out observations are obtained through the pooled information observed through observed measurements.

It is general tendency to analyze the trial data through the assumption of MAR. In MAR, the patients drop-out observations are obtained through the pooled information obtained through observed measurement.

The random effect model is useful tool for longitudinal data analysis. However, it is tedious work to handle the missing observation and compare the drug treatment effect.[14] The selection models and pattern mixture models are complex to apply in any longitudinal data. The interpretation and inference becomes cumbersome from the point of medical prospective. The simple additive model is required to apply with consideration of missing observation to increase the acceptability of drop-out in longitudinal data rather than ignore it.

Objective

The aim of this work is to compare the drug treatment effect of the type 2 diabetes patients observed in follow up visits. The biomarker serum cretinine has been considered in this work as response variable. The joint longitudinal analysis is applied with consideration of missing observation in the data set. The performance of combined drug therapy i.e, "Metformin with Pioglitazone" and "Gliclazide with Pioglitazone" has been compared to reducing the serum creatinine level. The joint longitudinal data models have been considered with MAR and MNAR assumption.

Modeling with Missing Observation

The semi-parametric Bayesian approach is effective to overcome the problem of missing repeated observations.[15] The weighted generalized estimating equation can also be considered as better estimation technique. In this connection, non-ignorable missing observations have been found with practice.[16] The non-ignorable missing observation can be fitted through pattern mixture modeling, selection modeling and shared-parameters modeling.[12,13] The estimating parameter becomes complex for the non-ignorable missing data. The selection approach modeled the missing data through conditional on complete data. The distribution of the missing data can be figured out through complete data in pattern mixture modeling. The random effect and dependence between measurements and missingness can only found in the shared-parameter models. Broadly, the handling of missing observation can be classified into four categories i.e. (I) Maximum likelihood (ML) (II) Multiple Imputation (III) Bayesian approach and (IV) weighted estimated equation. Here, the Bayesian approach has been applied to carry out the work.[15]

Method

From January 2007 to December 2008, 100 patients from Madurai are randomly selected in the clinical trial for comparing two different type 2 diabetes drug treatment effects. The standard therapy arm is consisted with "Metformin with Pioglitazone" drug combination. The experimental groups are consisted of "Pioglitazone with Gliclazide" drug combination. The drug therapies have been administered for a period of one year with three follow-up visits. Initially, a total of 123 patients are randomly selected in treatment and only 100 of them participate in the drug trial of which 50 patients are assigned to the standard therapy and rest to the experimental therapy. The outcomes are the overall control of sugar level and other biochemical parameters of the patients.

Assessment of Serum Creatinine

The joint modelling has been applied to compare the level of changes of serum creatinine. The measurements of serum creatinine have been observed for each patient in the three follow-up visits. The normal levels of serum creatinine are observed with 0.7 to 1.3 mg/dL for men and 0.6 to 1.1 mg/dL for women. The percentage of patients observed in each visits are provided in the Table 2.

A total of 100 patients have been observed in the data set, resulting in 80 (80%) complete cases. The lowest amount of missing observation has been observed at baseline 7(14%) with highest 10(20%) in the 3rd follow-ups. The information about death has not been found from any patients.

Joint Model with Missing at Random

The time effects have been considered as random. The p-value of smaller than 5% level is considered as significant. The relationship between serum creatinine and FBS, adjusting for sex, BMI, drug treatment effects has been studied under consideration of MAR. The imputed model is like,

$$serum\ creatinine_{ij} = \beta_{0,i,j} + \beta_1 FBS_{ij} + \beta_2 SEX_{ij} + \beta_3 BMI_{ij} + \beta_4 DRUG_{ij} \quad (1)$$

The relationship between serum creatinine and FBS, adjusting with sex, level of BMI and drug treatment effects has been studied.

Let i is the ID of the patients for the treatment j. The intercept term $\beta_{o,i,j}$ in equation (1) has been further separated to

$$\beta_{0,i,j}=\beta_0+\mu_j+e_{ij} \quad (2)$$

$$\mu_j \sim N(0,0.01) \quad (3)$$

$$e_{ij} \sim N(0,0.001) \quad (4)$$

The terms μ_j and e_{ij} have been assumed to follow the Normal distribution with minimum amount of standard deviation. The posterior mean of regression coefficient obtained through simulation procedures are given in the Table 1.

Joint Model with Missing Not at Random

In case of MNAR, the logistics regression has been applied in the equation (1). The variable serum creatinine has been classified into obs=1 or 0. If it is observed then given as,

logit $\Pr\{obs_{ij}=1\}$.

In both the case-MAR and MNAR , the imputed model as an extension of (1) in specified as,

$$FBS_{ij}=\beta_{0,i,j}^{imp}+ \beta_{0,i,j}^{imp}*\text{Serum creatinine}+ \beta_1^{imp}*\text{Sex}+ \beta_2^{imp}*\text{BMI}+ \beta_3^{imp}*\text{Drug} \quad (5)$$

$$\beta_{0,i,j}^{imp}=\beta_0^{imp}+\mu_j^{imp}+e_{ij}^{imp} \quad (6)$$

$$\mu_j^{imp}\sim N(0,0.01) \quad (7)$$

$$e_{ij}^{imp}\sim N(0,0.001) \quad (8)$$

The imputed model is like linear regression model, with the missing values of FBS imputed from the observed value of Serum Creatinine, Sex, BMI and Age respectively.

Statistical analysis

The Statistical analysis have been performed with R (Version 2.13.1) and WINBUGS 14. The cross sectional comparison of serum creatinine has been performed at each visits, through observed means and standard errors per visits. In this analysis, differences in serum creatinine level in each time point on both treatment groups are considered. The posterior means of the regression coefficients related to the level of response are computed through MCMC iterations. The initial 5,000 burns obtained through the MCMC techniques have been discarded. The standard deviation and 95% credible interval of each regression coefficient obtained through 5,000 times burns with 2 separate chains. The Highest Posteriors Density Intervals are also given in the Table for each regression coefficients. The mean value for the regression coefficients are shown in the Table 1. The uncertainty of the estimated regression parameter can be observed through the standard deviation value.

Discussion

The imputation technique has been applied in the type 2 diabetes drug treatment effect comparison. In recent years, the problem with missing observation with longitudinal analysis have received special attention.[18,19,20,21,22] In this paper, the additive model as a method to deal the drop-out observation has been applied. The methods have been illustrated on the randomized clinical trial data for patients with type 2 diabetes. The MAR and MNAR methods have been

compared and discussed. The Bayesian MCMC is useful to estimate unknown parameter in the joint modeling.[17] It is also useful in multivariate longitudinal and survival modeling to estimate the parameter with missing observation.[23] However, the trial with consideration of missing observations becomes difficult to interpret in presences of complex mathematical notation. In this scenario, the simple linear imputation technique has been applied in this work to compare the serum creatinine level among type 2 diabetes patients. The result shows that "Metformin with Pioglitazone" is better than "Pioglitazone with Gliclazide" to control the upper limit of serum creatinine over a period of one year study.

Conclusions

The presence of missing observation is natural in repeated measurement. The general tendency is to discard the missing cases before statistical analysis. The complete case analysis is the way to avoid the missing observation in the repeated measurement. However, the analysis with consideration of only fully observed cases reduces the power of the study. The other tendency is to overlook the trial having observation and conclusion with missing observation. This work is contributed to explore the serum creatinine as the drug treatment effect. The effect of fasting blood sugar has been found positively associated with serum creatinine. The reduction rate of serum after the end of the study among male and female are same. The result confirms that MAR and MNAR can be considered as powerful tool to overcome the missing observations in the clinical trial measurement. The covariates and prior observed measurement of the response variable assist to generate the unobserved information of the drop-out. It can be confirmed that "Metformin with Pioglitazone" is useful to control upper serum creatinine level than "Pioglitazone with Gliclazide". The elaborated method can be applied in other clinical trial problem to deal with missing observation in the follow-up study periods.

Conflict of Interest: None declared.

References:

1.Ferrannini E, Smith JD, Cobelli C, Toffolo G, Pilo A, DeFronzo RA. Effect of insulin on the distribution and disposition of glucose in man. *Journal of Clinical Investigation.* 1985; 76:357-364.

2.Yonemura K, Takahira R, Yonekawa O, Wada N, Hishida A: The diagnostic value of serum concentrations of 2-(alpha-mannopyranosyl)-Ltryptophan for normal renal function. *Kidney International.* 2004; 65:1395-1399.

3.Harita N, Hayashi T, Sato KK, Nakamura Y, Yoneda T, Endo G, et al.Lower serum creatinine is a new risk factor of type 2 diabetes: the Kansai healthcare study. *Diabetes Care* 2009; 32:424-426.

4.Chiong E, Wong AF, Chan YH, Chin CM. Review of clinical manifestations of biochemicallyadvanced prostate cancer cases. *Asian J Surg* 2005;28:202–206.

5.Vesalainen S, Lipponen P, Talja M, Syrjanen K. Biochemical parameters as prognostic factors in prostatic adenocarcinoma. *Acta Oncol* 1995;34:53–59.

6.Merseburger AS, Connelly RR, Sun L, Richter E, Moul JW. Use of serum creatinine to predict pathologic stage and recurrence among radical prostatectomy patients. *Urology* 2001;58:729–734.

7.Sandhu DP, Mayor PE, Sambrook PA, George NJ. Outcome and prognostic factors in patients with advanced prostate cancer and obstructive uropathy. *Br J Urol* 1992;70:412–416.

8.Ribeiro M, Ruff P, Falkson G. Low serum testosterone anda younger age predict for a poor outcome in metastatic prostate cancer. *Am J Clin Oncol* 1997;20:605–608.

9.Fitzmaurice GM, Laird NM and Ware JH. Applied longitudinal analysis. New York: *Wiley*; 2004.

10.Rubin D B. Inference and missing data. *Biometrika.*1976;63, 581–592.

11. Little RJA. Modeling the drop-out mechanism in repeated-measures studies. *Journal of the American Statistical Association*; 1995;90, 1112–1121.

12. Little RJA, Wang Y. Pattern-mixture models for multivariate incomplete data with covariates. *Biometrics* 1996;52:98–111.

13. Little RJA. Modeling the drop-out mechanism in repeated-measures studies. J *Am Stat Assoc.*1995;90:1113–1121.

14. Nath DC and Bhattacharjee A. Pattern Mixture Modeling. An Application in Anti-Diabetes Drug Therapy on Serum Creatinine on Type 2 Diabetes Patients. *Asian Journal of Mathematics and Statistics*.2012; 5(3);71-81.

15.Su L, Hogan JW. Bayesian semiparametric regression for longitudinal binary processes with missing data. *Statistics in Medicine*.2008; 27: 3247-3268.

16.Molenberghs G, Kenward MG, Lesaffre E. The analysis of longitudinal ordinal data with nonrandom drop-out. *Biometrika* 1997;84:33–34.

17. Faucett CL, Thomas DC. Simultaneously modelling censored survival data and repeatedly measured covariates: a Gibbs sampling approach. *Statistics in Medicine*,1996; 15(15):1663-85.

18. Daniels MJ, Hogan JW. Missing data in longitudinal studies. London: *Chapman and Hall*; 2008.

19. Verbeke, G, Molenberghs G. Linear mixed models for longitudinal data. New York: *Springer*; 2000.

20. Fitzmaurice GM.; Davidian M, Verbeke G,Molenberghs M. Longitudinal data analysis. London: *Chapman and Hall*; 2008.

21.Molenberghs G, Verbeke G. Models for discrete longitudinal data. New York: *Springer*; 2005.

22. Molenberghs G, Kenward MG. Missing data in clinical studies. New York: *Wiley*; 2007.

23. Chi YY, Ibrahim JG: Joint models for multivariate longitudinal and multivariate survival data. *Biometrics, 2006;* 62(2):432-45.

Table 1: Posterior mean (s.d.) for serum creatinine level at each time point, stratified by assumption of missing-ness ; MAR = missing at random; MNAR=Missing Not at random

	Parameter	Mean	SD	2.5%	97.5%
MAR	$B_{.FBS}$	0.03	0.06	0.00	0.08
	$B_{.DRUG}$	0.01	0.08	0.0001	0.0003
	$B_{.BMI}$	0.35	0.12	0.15	0.58
	B_0	0.59	0.14	0.35	0.83
	$\beta_{IMP.FBS}$	102.9	7.95	87.33	118.70
	$\beta_{IMP.DRUG}$	46.13	9.29	28.56	64.17
	$\beta_{IMP.BMI}$	29.47	8.78	12.85	47.27
	$\beta_{IMP.SEX}$	-3.45	6.34	-15.85	9.12
	$B_{0,IMP.}$	5.47	0.15	5.22	5.82
MNAR	$B_{.FBS}$	0.01	0.08	0.01	0.03
	$B_{.DRUG}$	0.39	0.15	0.16	0.67
	$B_{.BMI}$	0.35	0.12	0.15	0.58
	B_0	0.59	0.14	0.35	0.83
	$\beta_{IMP.FBS}$	102.9	7.95	87.33	118.70
	$\beta_{IMP.DRUG}$	46.13	9.29	28.56	64.17
	$\beta_{IMP.BMI}$	29.47	8.78	12.85	47.27
	$\beta_{IMP.SEX}$	-3.45	6.34	-15.85	9.12
	$B_{0,IMP.}$	5.47	0.15	5.22	5.82

Table 2: The Pattern of missingness in drug treatment group

Treatment	Observation	Number of missing components of Y_j	Percentage of missing observation
metformin with pioglitazone	srcret1st	0	0
	srcret 2nd	7	14
	srcret 3rd	8	16
gliclazide with pioglitazone	srcret1st	0	0
	srcret 2nd	4	8
	srcret 3rd	10	20