



Asian Journal of Mathematics & Statistics

ISSN 1994-5418

Pattern Mixture Modeling: An Application in Anti Diabetes Drug Therapy on Serum Creatinine in Type 2 Diabetes Patients

Dilip C. Nath and Atanu Bhattacharjee

Department of Statistics, Gauhati University, Guwahati-781014, India

Corresponding Author: Atanu Bhattacharjee, Department of Statistics, Gauhati University, Guwahati-781014, India

ABSTRACT

The handling of missing data is the difficult problem in longitudinal data analysis. There are broad ranges of methods to handle missing observation in longitudinal data. The comparison of different missing observation handling techniques by Last Observation Carry Forward (LOCF), EM algorithm, Missing at Random (MAR) and Missing not at Random (MNAR) with pattern mixture modeling has been applied on clinical trial data set, studying the effect of drug treatment of metformin with pioglitazone respect to pioglitazone with gliclazide in type 2 diabetes patients for three follow up. It has been found by using the MAR and MNAR approach, that pioglitazone with metformin is more effective to reduce the serum creatinine as compared to pioglitazone with gliclazide.

Key words: Dropout, pattern, multinomial, Bayesian

INTRODUCTION

India is burdened with type 2 diabetes and curable drug is emerging as one of the major therapy problems. Seyyednozadi *et al.* (2008) have shown the importance of physical activity to control the diabetes. Namazi *et al.* (2011) have found in a controlled clinical trial that hydro alcoholic extract of nettle has decreasing effect of TAC and SOD in type 2 diabetes patients. Sayyad *et al.* (2011) have demonstrated through computers simulation modeling for healthcare practice to preventing type 2 diabetes in India.

However, to better investigate the effect of treatment, one is often interested in evaluating how a biomarker of interest changes over time. This study is involved to provide clinical evidence in the search of effective treatment to the type 2 diabetes patients. The biomarker of interest has been considered with serum creatinine changes over the drug trial period.

UK Prospective Diabetes Study Group (1988) has confirmed that the metformin therapy alone has a 32% lower risk to develop any diabetes related complications in obese patients. Purnell and Hirsch (1997) have observed the effectiveness of metformin in non obese diabetic patients. Ertel (1998) have observed that pioglitazone group of drugs ameliorates for insulin resistance associated with Non-insulin-dependent Diabetes Mellitus (NIDDM) without stimulating insulin release.

The presence of missing observations in the data may be a reason for the inconsistent results. The reliability of the estimation goes down due to the high number of missing observation in the data set. In some cases, the missing observation could generate biased results in the inference on the response. Uysal (2007) has proposed an algorithm by radial basis function to deal with missing value problem in time series data analysis. Kayaalp and Cevyik (2001) have applied estimation of

missing observation in longitudinal data analysis to obtain the observation on water temperature. It has been found that the estimation through MCMC is very useful in this context. Suzuki *et al.* (2009) have applied the MCMC in the Linear Mixed modeling to estimating the optimal IBD vaccination timing in the DNA level. Jin *et al.* (2006) have developed the algorithm to estimate continuous-time speed data. It has also been concluded that the mentioned methods is useful in time-series modeling in presence of missing observations. Nath and Bhattacharjee (2011) have applied the Bayesian approach longitudinal data analysis in type 2 diabetes patients of India.

It is not unusual in studies for some sequence of measurement to terminate early for reason outside the control of this investigator and the affected unit is called as dropout. It might therefore be necessary to accommodate dropout in the modeling process to obtain correct inference for scientific interest. The preliminary work with missing observation has been done by Rubin (1976, 1987). They have presented an important distinction between different dropout values processes. To deal with the missing observations it is important to know the type of the missing observation present in the data, although it is difficult to understand the type of dropout. A dropout in the data is called as missing completely random (MCAR) if the dropout is independent of both observed data and unobserved data and Missing at Random (MAR) if, conditional on the observed data, the dropout is independent of unobserved measurement; otherwise it is called Missing at Non-random (MNAR). The likelihood approach can be applicable if the dropouts process is random and this situation is ignorable dropout Little and (Rubin, 1987). As Little (1993, 1994, 1995) has generalized the different class of models under the arena of Pattern-mixture Modeling (PMM). Statistical analysis with non-ignorable missing is based on the assumption of the missing data mechanism (Diggle and Kenward, 1994). Little (1993, 1994) has considered the PMM to stratify the incomplete data by the pattern of missing values and formulate the distinct models within each stratum. He has also presented a broad discussion on random-effects in the PMM to deal with the dropouts in 1995. Hogan and Laird (1997) have extended the PMM by permitting the censored dropout times, as might arise when there are late entrants to a trial and interim analyses are performed. Follmann and Wu (1995) and Wu and Bailey (1988) have used the conditional linear model to permit generalized linear models without any parametric assumption on the random effects model.

DeFronzo *et al.* (1981, 1985) have found that the skeletal muscle can be affected due to high amount of insulin secretion. Skeletal muscle is one of the major target organs due to the presences of metabolic creatinine in it (Zierath *et al.*, 2000). Martin (2003) has obtained that the skeletal muscle mass problem is associated with the type 2 diabetes. Currently there is insufficient evidence to suggest that a combined drug therapy protects the skeletal muscle mass in type 2 diabetic patients. Harita *et al.* (2009) have found the association of serum creatinine with type 2 diabetes. They have concluded that the low serum creatinine availability as the reason to develop type 2 diabetes in a study of non-obese middle-aged Japanese men. It has also been assessed whether a combined drug therapy could prevent the progression of serum creatinine secretions to protect skeletal muscle or not. The normal ranges of serum creatinine are from 0.44 to 0.83 mg dL⁻¹ for women and 0.56 to 1.23 mg dL⁻¹ for men (Yamamoto *et al.*, 2009). To the best of our knowledge, no clinical trials for type 2 diabetes drug therapies has been handled the missing observations with PMM.

The goal of this study is to compare the drug treatment effect by the low mean serum creatinine value during different visits. Different procedures have been applied to deal with missing observations in the data. The performance of combined drug therapy i.e., metformin with

pioglitazone and gliclazide with pioglitazone has been compared in reducing the serum creatinine level. Here, the missing observation in the serum creatinine of patients' at different visits have been handled by the LOCF, EM algorithm and pattern mixture modeling (i.e., MAR and MNAR).

MODEL SPECIFICATION

The Four different applied methods have been discussed below:

LOCF (last observation carry forward): In the clinical trial the missing data can be classified with non-informative (essentially random ;a study subject moves to another city) and informative (nonrandom, the study subject removed from the trial by the investigators). However, it is very much difficult to know whether data are non-informatively or informatively missing. The Last Observation Carried Forward (LOCF) is a commonly used way of imputing data with dropouts. Recently, Food and Drug Administration (FDA) has approved LOCF as the preferred method of analysis. It can be used as an imputation method when the data are longitudinal. The last observed value (non-missing value) is used to fill in missing values at a later point in the study.

EM algorithm: In case of EM algorithm (Dempster *et al.*, 1977), the imputed value Y can be obtained from the observed value of X by using the estimated value of $\hat{\alpha}_0$ and $\hat{\beta}_0$, i.e.:

$$Y_{\text{imp}} = \hat{\alpha}_0 + \hat{\beta}_0 X \tag{1}$$

However, the Eq. 1 is free with error term. It is based on the maximum likelihood estimates and converges to the population parameters.

The model can be called by:

$$Y_{\text{imp}} = \hat{\alpha}_{\text{EM}} + \hat{\beta}_{\text{EM}} X \tag{2}$$

where, $\hat{\alpha}_{\text{EM}}$, $\hat{\beta}_{\text{EM}}$ are the consistent estimated value of α_0 , β_0 obtained from final iterations.

Missing at random (MAR): In longitudinal studies to deal with the missing observations the widely used method is MAR, where the missing observations Y_{mis} are independent with the observed response and covariates in the model. Fitzmaurice *et al.* (2004) has proposed to use MAR as the default method to handle missing observation due to its independence with observed response and covariates.

In this context, Fitzmaurice *et al.* (2004) has also considered a model for the joint distribution of (Y_1, Y_2, S) , where $(Y_1, Y_2)^T$ is the bivariate response and S is a binary indicator of observed Y_2 . Fitzmaurice *et al.* (2004) has applied the PMM for the bivariate normal data by:

$$(Y_1, Y_2)^T | S=s \sim N(\mu^{(s)}, \sigma^{(s)}) \tag{3}$$

$$S \sim \text{Bernouli}(\phi) \tag{4}$$

where, for pattern $s \in \{0, 1\}$, the parameters are $\{\mu_1^{(s)}, \mu_2^{(s)}, \sigma_{11}^{(s)}, \sigma_{22}^{(s)}, \sigma_{12}^{(s)}\}$. We have reparameterized the model in Eq. 3, in term of the marginal distribution of Y_1 and the conditional distribution of Y_2 given Y_1 , such that:

$$Y_1|S = s \sim N(\mu_1^{(s)}, \sigma_{11}^{(s)}) \tag{5}$$

$$Y_2|Y_1, S = s \sim N(\beta_0^{(s)} + \beta_1^{(s)} Y_1, \sigma_{21}^{(s)}) \tag{6}$$

Here, we have assumed that the observed data and missing observation are in normal distribution and MAR respectively. The case $J = 3$ is applied due to presences of 3 observations of each patients. Let $(Y_1, Y_2, Y_3)^T$ denote the full data. In the model the dropout has been modeled in terms of number of observations, $S = \sum_j s_j$. The full data model has been factored as $P(Y_1, Y_2, Y_3, s) = P_s(Y_1, Y_2, Y_3) P(s)$. where S follows the multinomial distribution with $P_s = P(S = s)$.

The observed data response distributions within each pattern are specified using normal distributions:

$$Y_1|S = 1 \sim N(\mu^{(1)}, \sigma^{(1)}) \tag{7}$$

$$(Y_1, Y_2)^T|S = 2 \sim N(\mu^{(2)}, \sigma^{(2)}) \tag{8}$$

$$(Y_1, Y_2, Y_3)^T|S = 3 \sim N(\mu^{(3)}, \sigma^{(3)}) \tag{9}$$

Now consider identifying $P_s(y_{\text{mis}}|y_{\text{obs}})$ under MAR, Start with $S = 1$. We need to identify $P_1(y_2, y_3|y_1)$ which can be factored as:

$$P_1(y_2, y_3|y_1) = P_1(y_2|y_1) P_1(y_3|y_2, y_1) \tag{10}$$

The MAR constraints implies:

$$P_1(y_2|y_1) = P_{>2}(y_2|y_1) \tag{11}$$

$$P_1(y_3|y_2, y_1) = P_3(y_3|y_2, y_1) \tag{12}$$

The distribution $P_{>2}(y_2|y_1)$ can be called as the mixture Roy and Daniels (2008) of the normal distribution from pattern $S = 2$ and $S = 3$:

$$P_{>2}(y_2|y_1) = \frac{\phi_2 P_2(y_2|y_1) P_2(y_1) + \phi_3 P_3(y_2|y_1) P_3(y_1)}{\phi_2 P_2(y_1) + \phi_3 P_3(y_1)} \tag{13}$$

where, ϕ is the multinomial parameters with:

$$\sum_s \phi_s = 1, \phi_s = P(S = s); S \in 1, 2, 3 \text{ and } \phi = (\phi_1, \phi_2, \phi_3) \tag{14}$$

The model can be specified by the given Table 1.

In these models, we have applied the pattern by $S \in 1, 2, 3$. The presence of non-identified component has been denoted by *. The model is as follows:

$$Y_i|S = i \sim N(\mu^{(i)}, \sigma^{(i)}), i = 1, 2, 3 \tag{15}$$

Table 1: The non italic cell are fully observed data and italic cell are partially observed data of serum creatinine in different visits

	J = 1	J = 2	J = 3
S = 1	$P_1(Y_1)$	$P_1(Y_2/Y_1)$	$P_1(Y_3/Y_1, Y_2)$
S = 2	$P_2(Y_1)$	$P_{s,2}(Y_2/Y_1)$	$P_2(Y_3/Y_1, Y_2)$
S = 3	$P_3(Y_1)$	$P_{s,2}(Y_2/Y_1)$	$P_{s,3}(Y_3/Y_1, Y_2)$

$$*Y_2/Y_p, Y_1, S = 1 \sim N(\alpha_0^{(1)} + \alpha_1^{(1)}Y_1, \tau_2^{(1)}),$$

$$*Y_2/Y_p, Y_1, S = i \sim N(\alpha_0^{(i)} + \alpha_1^{(i)}Y_1, \tau_2^{(i)}), i = 1, 2, 3 \tag{16}$$

$$*Y_3/Y_1, Y_2, S = i \sim N(\beta_0^{(i)} + \beta_1^{(i)}Y_1 + \beta_2^{(i)}Y_2, \tau_2^{(i)}), i = 1, 2, 3 \tag{17}$$

$$S \sim \text{Mult}(\phi) \tag{18}$$

Here, β and τ are the regression parameter and variance component, respectively.

The model $p(y_{\text{mis}}/y_{\text{obs}}, s) p(y_{\text{obs}}/s) p(s)$ can be separated with $p(y_{\text{mis}}/y_{\text{obs}}, s)$, $p(y_{\text{obs}}/s)$ and $p(s)$.

To make out, $p_1(y_2/y_1)$, we have applied $p_1(y_2/Y_1)$ and:

$$\alpha_0^{(1)} + \alpha_1^{(1)} y_1 = \alpha_0^{(s2)} + \alpha_1^{(s2)} y_1, \tau_2^{(1)} = \tau_2^{(s2)} \tag{19}$$

all the possible values of y_1 , has been computed by, $\alpha_j^{(1)} = \alpha_j^{(s2)}$ for $j = 0, 1$.

The values of $p_1(y_3/y_1, y_2)$ and $p_2(y_3/y_1, y_2)$, has been obtained by solving the regression parameter and variance component of $\beta_i^s = \beta_0^s$ for $s = 1, 2$ and $i = 1, 2, 3$ and $\tau_{i=3}^s = \tau_3^s$.

The full data model is reparameterised to $\zeta = (\zeta_s, \zeta_M)$. Where, ζ_s has been generated from and ζ_s are elaborated to $\zeta_s = (\alpha_0^{(1)}, \alpha_1^{(1)}, \tau_2^{(1)}, \beta_0^{(s)}, \beta_1^{(s)}, \beta_2^{(s)}, S = 1, 2$ and $\zeta_M = (\mu^{(s)}, \sigma^{(s)}, \alpha_0^{(s2)}, \alpha_1^{(s2)}, \tau_2^{(s2)}, \beta_0^{(s)}, \beta_1^{(s)}, \beta_2^{(s)}, \tau_3^{(s)}, \phi_1, \phi_2, \phi_3$ for $s = 1, 2, 3$, respectively.

Here, ζ_s stands for full pattern of all observation and ζ_M for the missing observations. The distance from MAR has been obtained by ζ_s in place of ζ_M by a new parameter Δ with the function of $\zeta_s = h(\zeta_M, \Delta)$.

The process is as follows:

$$\begin{aligned} \alpha_0^{(1)} &= h_1(\zeta_M, \Delta) = \alpha_0^{(s2)} + \Delta_{\alpha_0}^{(12)} \\ \alpha_1^{(1)} &= h_2(\zeta_M, \Delta) = \alpha_1^{(s2)} + \Delta_{\alpha_1}^{(12)} \end{aligned} \tag{20}$$

Now, the reparametrization of β has been done by, $\beta_i^s = \beta_i^{(3)} + \Delta_{\beta_i}^{s3}$ with $s = 1, 2$ and $i = 1, 2, 3$.

The reparametrization of τ is denoted by, $\tau_2^{(1)} = \Delta_{\tau_2}^{(12)} \tau_2^{(22)}$, $\tau_2^{(2)} = \Delta_{\tau_2}^{(22)} \tau_2^{(22)}$, $\tau_3^{(1)} = \Delta_{\tau_3}^{(13)} \tau_3^{(3)}$ and:

$$\tau_3^{(1)} = \Delta_3^{(2:3)} \tau_3^{(3)} \tag{21}$$

The full set of Δ is represented by, $\Delta = (\Delta_\alpha, \Delta_\beta, \Delta_\tau)$.

Where:

$$\begin{aligned} \Delta_\alpha &= (\Delta_{\alpha_0}^{(12)}, \Delta_{\alpha_0}^{(22)}), \Delta_\beta = (\Delta_{\beta_0}^{(13)}, \Delta_{\beta_1}^{(13)}, \Delta_{\beta_2}^{(13)}, \Delta_{\beta_0}^{(23)}, \Delta_{\beta_1}^{(23)}, \Delta_{\beta_2}^{(23)}) \text{ and} \\ \Delta_\tau &= (\Delta_{\tau_2}^{(12)}, \Delta_{\tau_2}^{(22)}, \Delta_{\tau_3}^{(13)}) \end{aligned} \tag{22}$$

In this model, the Δ has been assumed as the hierarchical prior for the posterior distribution of ζ_M .

Missing not at random (MNAR): MNAR is the condition where the dropout is not related to the observed dependent variable. The MNAR can be expressed with the help of indicator variable. The mean value of serum creatinine obtained through the MNAR approach is based on the Pattern Mixture Modeling (PMM) procedure. The MNAR has been performed in the platform of MAR with the assumption of, $\Delta_\alpha = 0$, $\Delta_\beta = 0$ and $\Delta_\tau = 1$ (Daniels and Hogan, 2007).

The point is to be noted that in the PMM, the prior value of Δ does not influence the posterior mean value. Hogan and Daniels (2002) have separated the prior specification of Δ into the three parts by, $\Delta_{\alpha_0}^{(1,2)}$, $\Delta_{\beta_0}^{(1,2)}$ and $\Delta_{\beta_0}^{(2,3)}$.

Where:

$$\Delta_{\alpha_0}^{(1,2)} = E(Y_2 / Y_1, S=1) - E(Y_2 / Y_1, S \geq 2) \tag{23}$$

In this study, the $\Delta_{\beta_0}^{(1,2)}$ and $\Delta_{\beta_0}^{(2,3)}$, have been reformed to $\Delta_{\beta_0}^{(1,3)}$ to obtain the value of Δ by:

$$\Delta = (\Delta_{\alpha_0}^{(1,2)}, \Delta_{\beta_0}^{(1,3)}) \tag{24}$$

The subjective assumption is required to formulate the prior distribution of the parameters. In this analysis, Δ has been assumed to follow the uniform distribution. The variability of different approaches has been compared with the mean value changes of serum creatinine in different visits.

Application: The secondary data set has been obtained from a clinical trial in Menakshi Mission Hospital, Tamil Nadu in 2008. The data have been obtained by randomized, double blind, parallel group samples. The part of the data set consists of serum creatinine sample of 100 patients: 50 of them are grouped under treatment 1 (metformin with pioglitazone) and the rest of them are grouped under treatment 2 (gliclazide with pioglitazone).

This study is involved to model the response of serum creatinine on the type of therapies over a period of one year for various patients' profiles. Due to missing observations, difficulties have been faced in the response of the both groups. Patients are observed thrice with serum creatinine level during the 12-months study period.

Patient's meeting eligibility criteria are randomized to 1:1 by (a) metformin with pioglitazone and (b) gliclazide with pioglitazone, Fig. 1 shows a plot of serum creatinine changes over a period of time. It shows that in MNAR procedure the reductions of serum creatinine among the patients due to gliclazide with pioglitazone are faster in comparison to metformin with pioglitazone. In 1st month's visits, the gap between mean serum creatinine value of metformin with pioglitazone and gliclazide with pioglitazone has been increased after 12th month's visits. Subjects who have received the metformin with pioglitazone during the follow-up period are in higher serum creatinine from the study initiation to study end than subject with gliclazide with pioglitazone of type 2 diabetes.

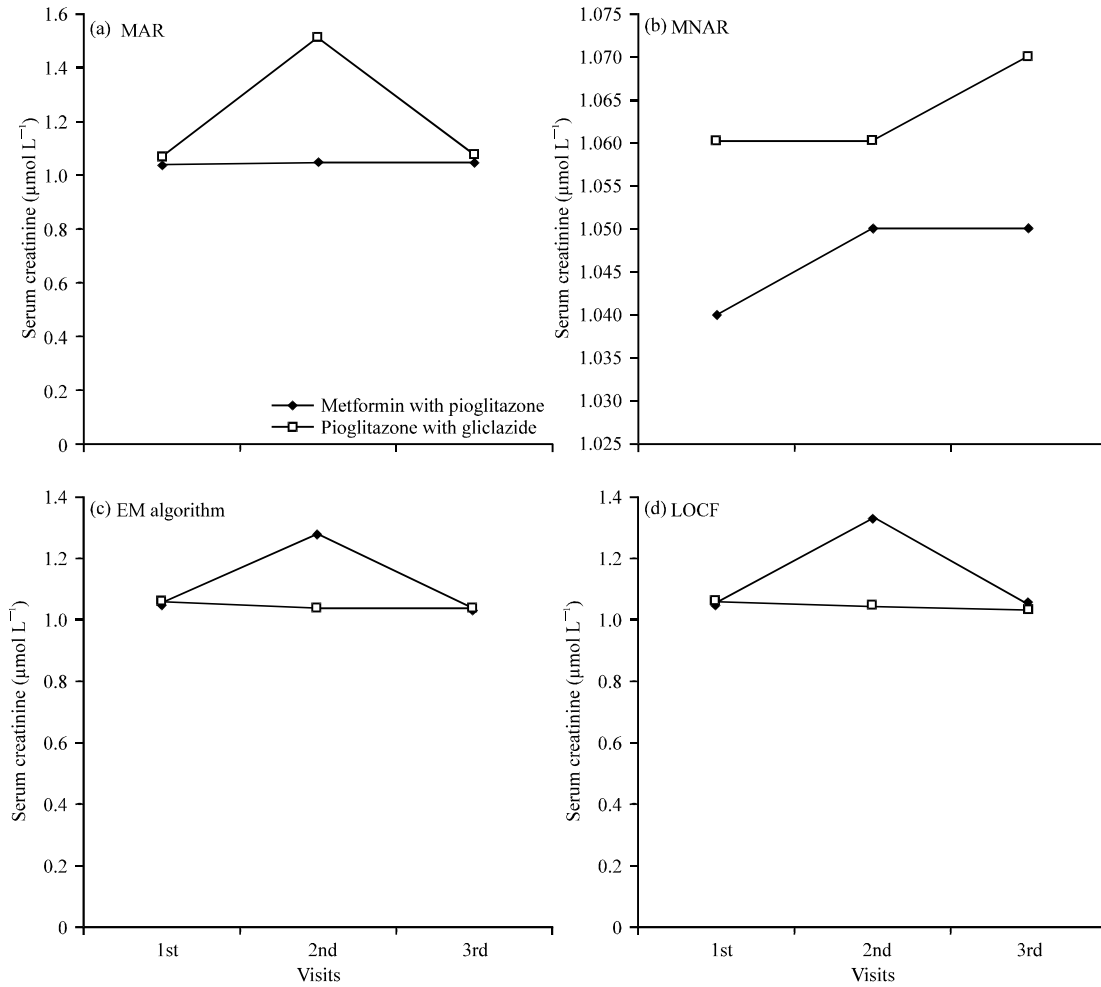


Fig. 1(a-d): Mean changes serum creatinine obtained in different drug groups with different techniques

Treatment options for type 2 diabetes in adolescents and youth have been randomized to parallel group trial consisting of a screening visit, a 2-6 month single-blind run-in period and a treatment period up to 12 months. The group under the therapy metformin with pioglitazone is observed with 14% (i.e., 7 out of 50) and 16% (i.e., 8 out of 50) missing observations in the 2nd and 3rd visits, respectively. The group under the therapy gliclazide with pioglitazone is observed by 8% (i.e., 4 out of 50) and 20% (i.e., 10 out of 50) missing observations in the 2nd and the 3rd visits, respectively. The observed data are displayed in Table 2. In the first stage, Expectation Maximization (EM) algorithm and LOCF have been applied, to deal with missing observations. In the second stage, MAR and MNAR assumptions have been carried out under the PMM approach.

Statistical analysis: The posterior means of serum creatinine in both the drug group has been obtained with the Monte Carlo Markov Chains (MCMC) of 20, 000 iterations using R2WINBUGS software. The first 5000 burn-in values have been discarded. The inferences of the mean value for both the therapy groups are given in Table 3. Uncertainty about Δ is reflected in the increased posterior standard deviation for parameters that rely on Δ (means at 2nd and 3rd) but importantly,

not for fully identified parameters (means at baseline). In this analysis, an informative prior for Δ and the posteriors are summarized in Table 3. The mean change has been obtained in different drug groups using different procedures are shown in Fig. 1.

RESULTS AND DISCUSSION

The numbers of patients that contribute to each visit are shown in Table 2. In both the drug groups the mean of serum creatinine has been reduced over the period of one year’s observation. In both the group there are large drop in the observed mean serum level at second visit. At the end of study, the means in both the groups are similar and remain rather constant and they are higher than at baseline. In the, 2nd and the 3rd visit the mean (sd) values have been obtained for the metformin with pioglitazone group through LOCF algorithm are 1.33 (1.75), 1.05 (0.12) compared to 1.28 (1.76) and 1.04 (0.13) by EM algorithm, respectively. However, the same result has been found in case of gliclazide with pioglitazone group by both the algorithm LOCF and EM (Table 2). The MNAR method produces the mean value at 1.05 (0.33) and 1.05 (0.96) in the 2nd and 3rd visits for the drug treatment group metformin with pioglitazone. The MNAR reveals the mean results for gliclazide with pioglitazone group by 1.05 (0.83) and 1.07 (0.62) for the subsequent visits. A small increase in the mean change is observed under imputation schemes of MAR and MNAR (Table 3). It can be concluded that MAR and MNAR produce same result with the different values of standard errors. The mean change in the (metformin with pioglitazone group is fairly robust to the regression imputation strategies, indicating similarities between the observed serum creatinine level i.e., metformin with pioglitazone group and gliclazide with pioglitazone group. However, in all the approaches, the absolute differences in mean serum creatinine after 1 year are small, as shown in Table 3.

Table 2: In different visits data availability description of serum creatinine

Treatment	Min.	Max.	Mean	NA observation	Available case
Metformin with pioglitazone					
Scret 1st	0.7	1.3	1.044	0	50
Scret 2nd	0.8	1.3	1.05	7	43
Scret 3rd	0.8	1.2	1.052	8	42
Gliclazide with pioglitazone					
Scret 1st	0.7	1.3	1.06	0	50
Scret 2nd	0.8	1.3	1.04	4	46
Scret 3rd	0.8	1.3	1.035	10	50

Table 3: Posterior mean (s.d.) for serum creatinine level at each time point, stratified by treatment group

Treatment	MAR	MNAR	EM	LOCF
Metformin with pioglitazone				
Scret 1st	1.04±0.15	1.04±0.63	1.05±0.14	1.05±0.14
Scret 2nd	1.05±0.14	1.05±0.33	1.28±1.76	1.33±1.75
Scret 3rd	1.05±0.15	1.05±0.96	1.04±0.13	1.05±0.12
Gliclazide with pioglitazone				
Scret1st	1.06±0.13	1.06±.38	1.06±0.16	1.06±0.16
Scret 2nd	1.51±0.15	1.05±0.83	1.04±0.13	1.04±0.13
Scret 3rd	1.07±0.16	1.07±0.62	1.04±0.13	1.03±0.14

Values are as Mean±SD. MAR: Missing at random, MNAR: Missing not at random, EM: Expected maximization algorithm, LOCF: Last observation carry forward

Methods to deal with the statistical information in type 2 diabetes drug therapy trials for which the biochemical parameters are missing is a neglected area in the literature. An observable solution is to overlook a trial with missing biochemical parameters. In this paper, our aim is to compare two treatment therapies for type 2 diabetes patient in South Indian population. This piece of work provides an application device to solve this problem by effectively predicting the missing observations that occurred in clinical trials. The pattern mixture model approach can be useful to compare drug treatment performance in different visits. In this study, it has been found by MAR and MNAR approach that metformin with pioglitazone is more effective to reduce the serum creatinine level as compared to gliclazide with pioglitazone. The presence of metformin is useful to reduce the level of serum creatinine. Aleisa *et al.* (2008) have confirmed that the metformin can be useful to avert the cardiac and hepatic toxicity in cancerous diabetes patients. The pattern mixture approach has been studied as a method for correction for informative drop-out. Analysis is complicated by presence of dropouts, since once a patient has been discontinued the treatment due to lost to follow up or clinical coordinator fails to observe, no longitudinal measurement of the covariates of interest can be collected thereafter. Thus, while analyzing such data, specific modeling is needed for appropriate and unbiased interpretations. Actually, PMM allows incorporating prior assumption for the drop out process in the drug treatment results. It does not differ much. First, a LOCF has been conventionally used to handle several types of drop outs to obtain the estimates of treatment effect directly related to the cumulative incidence of dropouts (Ambrogi *et al.*, 2008; Satagopan *et al.*, 2004). Secondly, an EM algorithm has been used for inference purposes (Williamson *et al.*, 2008). Actually, the Bayesian approach of Daniels *et al.* (2007) and Chi and Ibrahim (2006) has motivated our analysis of a Bayesian alternative that allows posterior inference for any parameter of interest. Thus, we have applied a fully Bayesian approach, implemented via MCMC methods using R2WinBUGS software (Faucett and Thomas, 1996; Chi and Ibrahim, 2006; Guo and Carlin, 2004; Ibrahim *et al.*, 2007). Such a Bayesian method for PMM modeling of longitudinal data is very recently reported (Hu *et al.*, 2009). We illustrated in this paper how the PMM approach may affect the results. We have compared the results of the pattern mixture approach with the results of the commonly used LOCF and EM algorithm. The LOCF and EM algorithm have produced inverse results in comparison to MAR and MNAR approach. Present results suggest that the treatment effect of the metformin with pioglitazone is effective to reduce more serum creatinine level as compared to gliclazide with pioglitazone.

CONCLUSIONS

In the study, we have presented a novel model to obtain serum creatinine level in type 2 diabetes patients and, therefore, the drug effect comparison. Present results confirm that MAR and MNAR are powerful tools for longitudinal data analysis and, consequently, for the actual application to the drug effect comparison in clinical trial. We have employed Markov Chain Monte Carlo methods to effectively estimate Serum Creatinine values for different visits in type 2 diabetes patients. The missing observation techniques have been applied in the drug treatment effect comparison in clinical trial data. We believe that more research is needed in this area. On the basis of the results reached we think the model can be applicable to area like Oncology, Aids and in other disease specific clinical trial. The combination of metformin with pioglitazone is more effective to reduce the serum creatinine and hence the risk of type 2 diabetes. The MAR and MNAR approach attaches the efforts of how the trial could change our opinion about the treatment effect. The statistical models with prior information need to be considered regarding information on lost observations in drug treatment effect comparison.

REFERENCES

- Aleisa, A.M., S.S. Al-Rejaie, S.A. Bakheet, A.M. Al-Bekairi and O.A. Al-Shabanah *et al.*, 2008. Protective effect of metformin on cardiac and hepatic toxicity induced by adriamycin in swiss albino mice. *Asian J. Biochem.*, 3: 99-108.
- Ambrogi, F., E. Biganzoli and P. Boracchi, 2008. Estimates of clinically useful measures in competing risks survival analysis. *Stat. Med.*, 27: 6407-6425.
- Chi, Y.Y. and J.G. Ibrahim, 2006. Joint models for multivariate longitudinal and multivariate survival data. *Biometrics*, 62: 432-445.
- Daniels, M.J. and J.W. Hogan, 2007. *Missing Data in Longitudinal Studies: Strategies for Bayesian Modeling and Sensitivity Analysis*. 1st Edn., Chapman and Hall/CRC, USA, ISBN-13: 978-1584886099, Pages: 328.
- DeFronzo, R.A., E. Jacot, E. Jequier, E. Maeder, J. Wahren and J.P. Felber, 1981. The effect of insulin on the disposal of intravenous glucose: Results from indirect calorimetry and hepatic and femoral venous catheterization. *Diabetes*, 30: 1000-1007.
- DeFronzo, R.A., R. Gunnarsson, O. Bjorkman, M. Olsson and J. Wahren, 1985. Effects of insulin on peripheral and splanchnic glucose metabolism in noninsulin-dependent (type II) diabetes mellitus. *J. Clin. Invest.*, 76: 149-155.
- Dempster, A.P., N.M. Laird and D.B. Rubin, 1977. Maximum likelihood from incomplete data via the EM algorithm. *J. R. Stat. Soc.*, 39: 1-38.
- Diggle, P. and M.G. Kenward, 1994. Informative dropout in longitudinal data analysis. *Applied Stat.*, 43: 49-93.
- Ertel, N.H., 1998. New therapeutic agents for type 2 diabetes: In managing type 2 diabetes. *Postgraduate Medicine Special report March*, 25-32
- Faucett, C.L. and D.C. Thomas, 1996. Simultaneously modelling censored survival data and repeatedly measured covariates: A Gibbs sampling approach. *Stat. Med.*, 15: 1663-1685.
- Fitzmaurice, G.M., N.M. Laird and J.H. Ware, 2004. *Applied Longitudinal Analysis*. John Wiley and Sons Inc., New York.
- Follmann, D. and M.C. Wu, 1995. An approximate generalized linear model with random effects for informative missing data. *Biometrics*, 51: 151-168.
- Guo, X. and B.P. Carlin, 2004. Separate and joint modeling of longitudinal and event time data using standard computer packages. *Am. Stat.*, 58: 16-24.
- Harita, N., T. Hayashi, K.K. Sato, Y. Nakamura, T. Yoneda, G. Endo and H. Kambe, 2009. Lower serum creatinine is a new risk factor of type 2 diabetes: The Kansai healthcare study. *Diabetes Care*, 32: 424-426.
- Hogan, J.W. and N.M. Laird, 1997. Mixture models for the joint distribution of repeated measures and event times. *Stat. Med.*, 16: 239-257.
- Hogan, J.W. and M.J. Daniels, 2002. A hierarchical modelling approach to analyzing longitudinal data with drop-out and non-compliance, with application to an equivalence trial in paediatric acquired immune deficiency syndrome. *J. Royal Stat. Soc.*, 51: 1-21.
- Hu, W., G. Li and N. Li, 2009. A Bayesian approach to joint analysis of longitudinal measurements and competing risks failure time data. *Stat. Med.*, 28: 1601-1619.
- Ibrahim, J.G., M. Chen and D. Sinha, 2007. Bayesian methods for joint modeling of longitudinal and survival data with application to cancer vaccine trials. *Stat. Sinica*, 14: 863-883.
- Jin, X., S.K. Hong and Q. Ma, 2006. An algorithm to estimate continuous-time traffic speed using multiple regression model. *Inform. Technol. J.*, 5: 281-284.

- Kayaalp, G.T. and F. Cevyk, 2001. Estimation of missing values in longitudinal data sets using regression methods in biological research. *J. Biol. Sci.*, 1: 678-679.
- Little, R.J.A., 1993. Pattern-mixture models for multivariate incomplete data. *J. Am. Stat. Assoc.*, 88: 125-133.
- Little, R.J.A., 1994. A class of pattern-mixture models for normal incomplete data. *Biometrika*, 81: 471-483.
- Little, R.J.A., 1995. Modeling the dropout mechanism in repeated-measures studies. *J. Am. Stat. Assoc.*, 90: 1112-1121.
- Martin, R.F., 2003. Renal Function. In: *Clinical Chemistry Theory, Analysis, Correlation*, Kaplan, L.A., A.J. Pesce and S.C. Kazmierczak (Eds.). St. Louis, Missouri, Mosby, pp: 483-484.
- Namazi, N., A.T. Esfanjani, J. Heshmati and A. Bahrami, 2011. The effect of hydro alcoholic nettle (*Urtica dioica*) extracts on insulin sensitivity and some inflammatory indicators in patients with type 2 diabetes: A randomized double-blind control trial. *Pak. J. Biol. Sci.*, 14: 775-779.
- Nath, D.C. and A. Bhattacharjee, 2011. A bayesian approach in autoregressive models in longitudinal data analysis: An application to type 2 diabetes drug comparison. *Asian J. Applied Sci.*, 4: 640-648.
- Purnell, J.Q. and I.B. Hirsch, 1997. New oral therapies for type 2 diabetes. *Am. Fam. Phys.*, 56: 1835-1842.
- Roy, J. and M.J. Daniels, 2008. A general class of pattern mixture models for nonignorable dropout with many possible dropout times. *Biometrics*, 64: 538-545.
- Rubin, D.B., 1987. *Multiple Imputation for Nonresponse in Surveys* John Wiley and Sons, New York.
- Rubin, D.B., 1976. Inference and missing data. *Biometrika*, 63: 581-592.
- Satagopan, J., L. Ben-Borat, M. Berwick, M. Robson, D. Kutler and A. Auerbach, 2004. A note on competing risks in survival data analysis. *Br. J. Cancer*, 91: 1229-1235.
- Sayyad, M.G., A.K. Shahani, S. Pal, J.A. Shirodkar and G. Gopal, 2011. Computer simulation modelling for the prevention of diabetes mellitus (prameha). *J. Applied Sci.*, 11: 2670-2679.
- Seyyednozadi, M., M.T. Shakeri, R. Rajabian and A. Vafaei, 2008. Role of physical activity and nutrition in controlling type 2 diabetes mellitus-2007. *J. Boil. Sci.*, 8: 794-798.
- Suzuki, K., J. Caballero, F. Alvarez, M. Faccioli, M. Goreti, M. Herrero, M. Petruccioli, 2009. The use of linear mixed models to estimate optimal vaccination timing for infectious bursal disease in broilers in paraguay. *Int. J. Poult. Sci.*, 8: 363-367.
- UK Prospective Diabetes Study Group, 1988. Effect of intensive blood glucose control with metformin on complications in overweight patients with type II diabetes. *Lancet*, 352: 854-865.
- Uysal, M., 2007. Reconstruction of time series data with missing values. *J. Applied Sci.*, 7: 922-925.
- Williamson, P.R., R. Kolamunnage-Dona, P. Philipson and A.G. Marson, 2008. Joint modelling of longitudinal and competing risks data. *Stat. Med.*, 27: 6426-6438.
- Wu, M.C. and K.R. Bailey, 1988. Analysing changes in the presence of informative right censoring caused by death and withdrawal. *Stat. Med.*, 7: 337-346.
- Yamamoto, M., M. Yamauchi, T. Yamauchi and T. Sugimoto, 2009. Low serum level of the endogenous secretory receptor for advanced glycation end products (esRAGE) is a risk factor for prevalent vertebral fractures independent of bone mineral density in patients with type 2 diabetes. *Diabetes Care*, 32: 2263-2268.
- Zierath, J.R., A. Krook and H. Wallberg-Henriksson, 2000. Insulin action and insulin resistance in human skeletal muscle. *Diabetologia*, 43: 821-835.