Published in final edited form as:

J Appl Stat. 2012 ; 39(3): 631–642. doi:10.1080/02664763.2011.610441.

Slope Estimation of Covariates that Influence Renal Outcome following Renal Transplant Adjusting for Informative Right Censoring

Miran A. Jaffa^{1,#}, Ayad A Jaffa^{2,3}, and Stuart R. Lipsitz⁴

¹Epidemiology and Population Health Department, Faculty of Health Sciences, American University of Beirut, Beirut, Lebanon

²Department of Biochemistry, American University of Beirut, Beirut, Lebanon

³Department of Medicine, Medical University of South Carolina, Charleston, SC. 29425, USA

⁴Harvard Medical School, Boston, Massachusetts, 02120, USA

Abstract

A new statistical model is proposed to estimate population and individual slopes that are adjusted for covariates and informative right censoring. Individual slopes are assumed to have a mean that depends on the population slope for the covariates. The number of observations for each individual is modeled as a truncated discrete distribution with mean dependent on the individual subjects' slopes. Our simulation study results indicated that the associated bias and mean squared errors for the proposed model were comparable to those associated with the model that only adjusts for informative right censoring. The proposed model was illustrated using renal transplant dataset to estimate population slopes for covariates that could impact the outcome of renal function following renal transplantation.

Keywords

Discrete geometric distribution; Empirical Bayes estimates; Informative right censoring; Longitudinal data; likelihood function; Slope estimation

1. Introduction

Longitudinal studies involving repeated data measurements for the same subjects over time are of great interest in the biomedical field. Since such data enable researchers to assess and monitor the rate of change in the outcomes of interest over a period of time. However, this type of data, although very informative to researchers, is also very prone to having missing values that could be due to subject withdrawal from the study. There are essentially two types of missing data, one that is missing at random and one that is not and is referred to as informatively missing. Informative right censoring is one type of missing process that occurs

[#]Correspondence: Miran A. Jaffa., Epidemiology and Population Health Department, Faculty of Health Sciences, American University of Beirut, Beirut, Lebanon, ms148@aub.edu.lb.

Conflict of interest: None declared.

when the dropout from the study is related to the subject's true rate of change known as slopes. In this context, ignoring the informative censoring or treating the missing data as randomly missing may generate inaccurate estimates of the rate of change for the outcomes of interest. A typical example of longitudinal studies that suffer from informative right censoring is the cohort of renal transplantation. This patient population is prone to graft rejection and renal failure following transplantation which leads to patients' withdrawal from the study. Accordingly, complete set of repeated measurements on biomarkers of their renal function can no longer be attained. In addition renal outcome was shown to be influenced by covariates such as race and gender of donors and recipients [11, 12]. In order to estimate population slopes for these covariates, a previous study employed the model that only adjusts for informative right censoring on groups of patients classified together by race and gender [3]. However, such an approach may result in groups with small sample size and potential loss of statistical power.

To estimate slopes that are adjusted for informative right censoring several statistical approaches were presented. Among these approaches were weighted and unweighted least square estimates [6, 17], the efficient non-iterative approach of the random growth curve models with unbalanced data [16], in addition to the iterative approaches that use the restricted or maximum likelihood estimates [4,7,8]. However, all these statistical approaches assume that the missing mechanism occurs in a random fashion [5]. Assuming that the missing process is non-ignorable when in fact it is may result in invalid biased estimates [18, 19]. In particular, if informative right censoring is ignored then this causes bias in standard likelihood-based analyses, as well as in weighted averages of individual least squares slopes [14].

Different approaches have been proposed to provide a valid estimate for population slopes in the presence of informative right censoring. A statistical method that assumes linear random effects model in the presence of informative right censoring has been previously developed using probit pseudo-maximum likelihood estimator (PPMLE) of the population slope [19]. In this approach, it was assumed that the censoring is governed by a probit function of both initial value and the slope of the individual subjects. An alternative approach adapted a conditional linear model to develop a linear minimum variance unbiased (LMVUB) and linear minimum mean squared error (LMMSE) estimate of the population slope [18]. A lognormal survival model was also proposed to provide a maximum likelihood estimate of the population slope using an EM algorithm [14]. All these methods focused only on population slope and did not provide an estimate for the individual subjects' slopes. One approach to estimate the population as well as individual slopes was proposed [9]. In this model the number of observations for each subject was modeled and was assumed to follow a geometric distribution with mean dependent on the individual slope. In their study [9], maximizing the likelihood function required solving a set of complicated equations iteratively for the set of parameters needed to be estimated using adjustment methods proposed by Stefanski and Carroll [15]. Iteration between the different equations is initiated till convergence. This approach tends to be computationally intensive and successful convergence may not be always achieved as noted by the authors. In addition, this model does not adjust for the effect of covariates when estimating the population and individual slopes.

In this study we developed a model that estimates population and individual slopes that are adjusted for covariates and informative right censoring. Specifically, we propose a likelihood-based approach where direct maximization of the approximate likelihood function integrated over the random effect is employed using adaptive Gaussian quadrature method [10]. This approach of direct maximization of the marginal likelihood function resulted in successful convergence for the parameter estimates. Here, individual slopes are assumed to be normally distributed with mean dependent on the covariates X_i and their corresponding population slopes θ_1 . The number of observations for each individual is assumed to follow a discrete distribution that depends on the censoring parameter γ_1 and the individual subjects slopes β_i . Consequently the number of observations depends on the covariate X_i and the population slope θ_1 for this covariate via the individual subject slopes β_i .

2. The Model

Consider a set of $i=1, \dots, n$ independent subjects, with outcome y_{ik} at measurement times $k=(t_{i1}, t_{i2}, \dots, t_{im_i})$ where m_i is the number of observations for subject i ($2 \leq m_i \leq p$). Every subject is assumed to have at least 2 observations so that slopes can be estimated and not to exceed P the number of prespecified measurement time points. The underlying model for y_{ik} is assumed to be

$$y_{ik} = \alpha_i + \beta_i t_{ik} + e_{ik}, \quad (1)$$

with random effects α_i and β_i , and random error e_{ik} , which is assumed to be normal with mean zero and variance σ_ε^2 . Given α_i and β_i , y_{ik} is normal with mean $\alpha_i + \beta_i t_{ik}$ and variance σ_ε^2 . Ordinary least squares (OLS) estimate of β_i are,

$$b_{i,ols} = \frac{\sum_{k=1}^{m_i} (t_{ik} - \bar{t}_i) y_{ik}}{\sum_{k=1}^{m_i} (t_{ik} - \bar{t}_i)^2}, \quad (2)$$

where $\bar{t}_i$ is the sample mean of the observation times for subject i . Further, given α_i and β_i ,

$$\text{Var}(b_{i,ols}) = \sigma_{b_i}^2 = \frac{\sigma_\varepsilon^2}{\sum_{k=1}^{m_i} (t_{ik} - \bar{t}_i)^2} \quad (3)$$

Thus, for each individual subject the following model is assumed:

1.

$$\beta_i \sim \text{Normal}(\theta_0 + \theta_1 X_i, \sigma_\beta^2); \quad (4)$$

2.

$$m_i \beta_i \sim \text{Truncated Discrete Distribution with } \Pr(M=m_i) = F(\gamma_0 + \gamma_1 \beta_i); \quad (5)$$

3.

$$b_{i,ols}|m_i, \beta_i \sim \text{Normal}(\beta_i, \sigma_{b_i}^2). \quad (6)$$

The underlying individual subjects' slopes β_i 's are normally distributed with a mean $\theta_0 + \theta_1 X_i$ and variance σ_β^2 (assumption 1). Population slopes θ_1 correspond to the vector of covariates X_i . Conditional on the individual slopes β_i 's, the number of observations for each individual m_i is modeled and assumed to follow a discrete distribution with probability dependent on the individual's true rate of change β_i and thus implicitly on the covariates X_i (assumption 2). Any discrete distribution that fits the dataset under consideration can be modeled. The censoring parameters are denoted as γ_0 and γ_1 , with $\gamma_1 = 0$ if censoring is informative and equals to zero otherwise. The discrete distribution is left-truncated since for each individual subject at least two observations are required to estimate individual slopes, and right-truncated since the study may terminate before observing the withdrawal of every subject from the study. The OLS estimates for individual subjects $b_{i,ols}$ ($i = 1, 2, \dots, n$) follow a normal distribution with mean β_i and variance $\sigma_{b_i}^2$ and inherently depend on X_i (assumption 3). The likelihood function corresponding to this model can be described as follows:

$$\begin{aligned} L &= \prod_{i=1}^n f(\beta_i, m_i, b_{i,ols}) = \prod_{i=1}^n f(b_{i,ols}|m_i, \beta_i) * f(m_i, \beta_i) \Rightarrow \\ L &= \prod_{i=1}^n f(\beta_i) f(m_i|\beta_i) f(b_{i,ols}|m_i, \beta_i) \end{aligned} \quad (7)$$

This likelihood function (eq.7) will then be integrated over the unobserved β_i to get a marginal likelihood,

$$L = L(\theta_0, \theta_1, \gamma_0, \gamma_1) = \prod_{i=1}^n \int f(m_i, b_{i,ols}) f(\beta_i) d\beta_i = \prod_{i=1}^n \int f(\beta_i) f(m_i|\beta_i) f(b_{i,ols}|\beta_i, m_i) d\beta_i \quad (8)$$

which will be directly maximized to obtain maximum likelihood estimate (MLE) of the population slopes θ_1 , intercepts θ_0 and censoring parameters (γ_0, γ_1) . Moreover, individual random slopes β_i are predicted using Empirical Bayes method described in Pinheiro and Bates [10]. Note that the likelihood is also a function of the unknown parameters σ_ε^2 which is the unknown parameter in $\text{Var}(b_{i,ols})$ and σ_β^2 . Conditional on α_i and β_i in equation 1, the unknown parameter σ_ε^2 in the likelihood function (eq.8), will be replaced by s_ε^2 the unbiased OLS estimate of σ_ε^2

$$s_\varepsilon^2 = \frac{\sum_{i=1}^n \sum_{k=1}^{m_i} (y_{ik} - \bar{y}_i - b_i(t_{ik} - \bar{t}_i))^2}{\sum_{i=1}^n m_i - 2n}, \quad (9)$$

The marginal likelihood function (eq.8) which encloses s_ε^2 (eq.9) will directly be maximized to obtain consistent estimates of the other parameters of interest $(\theta_0, \theta_1, \gamma_0, \gamma_1, \sigma_\beta^2)$. Direct

maximization of the likelihood function was conducted following dual Quasi-Newton optimization algorithm [1] using statistical SAS procedure NLMIXED [13].

3. Discrete geometric model

Truncated discrete geometric distribution is assumed for the number of observations but any discrete distribution could be assumed instead. In particular, in assumption 2 we consider the following:

2) $m_i | \beta_i \sim \text{Truncated geometric with } P_i = F(\gamma_0 + \gamma_1 \beta_i);$

$$\text{with } \Pr(M=m_i) = \{1 - F(\gamma_0 + \gamma_1 \beta_i)\}^{m_i-2} \{F(\gamma_0 + \gamma_1 \beta_i)\}^{R_i}, m_i=2, 3, \dots, p \quad (10)$$

$$\text{and } F(\gamma_0 + \gamma_1 \beta_i) = \frac{1}{1 + \exp[-(\gamma_0 + \gamma_1 \beta_i)]}$$

As was previously defined [9], R_i is an indicator variable that takes the value 1 if censoring occurred and zero otherwise. The geometric random variable is described as being the number of Bernoulli trials needed to achieve the first success that is in our case the censoring event. However, the geometric distribution used in this study is not the typical geometric distribution whereby success is always observed. The indicator variable R_i introduced to the probability function was originally proposed by Mori et al [9] to account for earlier termination of the study before observing withdrawal of all subjects. If withdrawal occurs during the period of the study then $R_i = 1$ and thus $m_i < p$. Note that m_i is the number of observations for every individual and $2 \leq m_i \leq p$. However if withdrawal does not occur during the course of the study then the subject has a complete set of observations with $m_i = p$ and $R_i = 0$ and this causes the right truncation of the distribution. Traditional geometric distribution assumes that success (withdrawal in our case) is going to occur and is observed. If the traditional geometric distribution is used here without the R_i then the probability model would be $\Pr(M=m_i) = \{1 - F(\gamma_0 + \gamma_1 \beta_i)\}^{m_i-2} F(\gamma_0 + \gamma_1 \beta_i)$. In this case right truncation is not accounted for since we are assuming that all subjects withdrew before the end of the study. However, by introducing R_i , right truncation was accommodated for by having $R_i = 0$ (that is no withdrawal has occurred which results in the right truncation of the distribution). The corresponding probability model differs from the traditional geometric distribution and will therefore be

$$\Pr(M=m_i) = \{1 - F(\gamma_0 + \gamma_1 \beta_i)\}^{m_i-2} = \{1 - F(\gamma_0 + \gamma_1 \beta_i)\}^{p-2}$$

For those who withdrew from the study, success was observed and $R_i = 1$. This withdrawal does not cause the right truncation of the model. In this case, the corresponding probability model would be the same as the traditional geometric distribution with probability function

$$\Pr(M=m_i) = \{1 - F(\gamma_0 + \gamma_1 \beta_i)\}^{m_i-2} F(\gamma_0 + \gamma_1 \beta_i).$$

Discrete geometric model was shown to be applicable to situations when the risk of withdrawal is high at the beginning but then decreases with time [9]. Accordingly this model suits well studies on organ transplants, one of which is the renal transplantation. In this setting, the risk of graft rejection is at its peak just after the transplantation since the body is more likely to refuse the new organ. However, this risk tends to decline with time.

The likelihood function corresponding to the discrete geometric distribution model is defined as follows:

$$\begin{aligned}
 L &= L(\theta_0, \theta_1, \gamma_0, \gamma_1) = \prod_{i=1}^n f(m_i, b_{i,ols}) f(\beta_i) d\beta_i = \prod_{i=1}^n f(\beta_i) f(m_i|\beta_i) f(b_{i,ols}|\beta_i, m_i) d\beta_i \\
 &= \prod_{i=1}^n \int \frac{1}{\sqrt{2\pi}\sigma_\beta} \exp \left\{ -\frac{1}{2} \left(\frac{\beta_i - (\theta_0 + \theta_1 X_i)}{\sigma_\beta} \right)^2 \right\} * \{1 - F(\gamma_0 + \gamma_1 \beta_i)\}^{m_i-2} * \{F(\gamma_0 + \gamma_1 \beta_i)\}^{R_i} \\
 &\quad * \frac{1}{\sqrt{2\pi}\sigma_{b_i}} \exp \left\{ -\frac{1}{2} \left(\frac{b_{i,ols} - \beta_i}{\sigma_{b_i}} \right)^2 \right\} d\beta_i
 \end{aligned} \quad (11)$$

The log of this likelihood function was maximized using SAS procedure NLMIXED and estimates of the individual slopes β_i , the population slope θ_1 corresponding to the covariate of interest X_i , intercept θ_0 , censoring parameters γ_0 and γ_1 , and σ_β^2 were all obtained.

4. Simulation Study

A simulation study was conducted to assess the performance of the model that adjusts for covariates and informative right censoring. Datasets were simulated such that the number of observations follows a discrete geometric distribution with

$$\Pr(M=m_i) = \{1 - F(\gamma_0 + \gamma_1 \beta_i)\}^{m_i-2} \{F(\gamma_0 + \gamma_1 \beta_i)\}^{R_i}, m_i=2, 3, \dots, 7$$

and the values for the covariate X_i were simulated from a Bernoulli distribution with a probability of success π , i.e., $\Pr(X_i = 1) = \pi$. The effect of the value of π on the performance of the model with binary covariate was assessed by incorporating different values for the probability of success. The number of observations ranged from 2 to 7 recorded at prespecified time points (t_{ij}) 0, 1, 3, 6, 12, 24, and 36 months. A linear relationship was assumed between the outcome and $\log(t_{ij}+1)$ such that $y_{ij} = \alpha_i + \beta_i \log(t_{ij}+1) + e_{ij}$. The errors e_{ij} were assumed to be normally distributed with mean zero and variance σ_e^2 . Estimates for the error variance (eq. 9) were used in the likelihood function. The performance of the model adjusting for covariates was assessed by bias, mean squared error for the population slope (MSE(a)) and mean squared error for the individual slopes (MSE(b)). MSE(a) was used to assess the performance of the model when the population slope is estimated and was defined as follows [9]:

$$MSE(a) = \frac{1}{r} \sum_{s=1}^r (\hat{\theta}_s - \theta_1)^2 \quad (12)$$

with r being the total number of replications. Meanwhile, $MSE(b)$ was used to assess the performance of the proposed model when the individual subjects' slopes were estimated and was defined as follows [9]:

$$MSE(b) = \frac{1}{nr} \sum_{s=1}^r \sum_{i=1}^n (\hat{\beta}_{is} - \beta_{is})^2 \quad (13)$$

with n being the size of the data set generated using Monte Carlo algorithm for simulation. In Table 1, we present the results of the simulation study when the number of observations were simulated from the geometric distribution and the covariate X_i was simulated from a Bernoulli distribution with different values of probability of success: 1) $\Pr(X_i = 1) = \pi = 0$ 2) $\Pr(X_i = 1) = \pi = 0.25$, 3) $\Pr(X_i = 1) = \pi = 0.5$ and 4) $\Pr(X_i = 1) = \pi = 0.75$. The performance of the model with a continuous covariate was also examined whereby we simulated covariate X_i from a normal distribution with mean 43.51 and standard deviation of 13.9. These values represent the mean and standard deviation of the age of the patients in the renal transplant dataset. Different censoring parameters were incorporated in the simulation to assess the effect of the censoring level on the performance of the geometric model on the normally distributed covariates and on the different values of the probability of success for binary covariates. In particular we incorporated no to low censoring levels indicated as $\gamma_1=0$ and $\gamma_1=2.3$ and medium to high level censoring with $\gamma_1=4.5$ and $\gamma_1=6.2$. Datasets simulated with no censoring to low-level censoring experienced about 14% of informative censoring, meanwhile those simulated with medium to high level censoring had about 17% informative censoring.

5. Discussion of the simulation results

We proceed first by discussing the effect of the different censoring levels at the various values of probability of success. In this context, we realized that when the censoring was non-informative, i.e., $\gamma_1 = 0$, bias associated with estimating the population slope, $MSE(a)$ and $MSE(b)$ across all the probability levels, were about -0.073, 0.018 and 0.013 respectively. When the censoring level was low, i.e., $\gamma_1 = 2.3$, the overall bias was -0.074, the $MSE(a)$ was 0.016 and $MSE(b)$ was 0.012. However, at medium and high censoring levels with $\gamma_1 = 4.5$ and $\gamma_1 = 6.2$ respectively the average bias was about -0.077, $MSE(a)$ was about 0.016 and $MSE(b)$ was 0.012. Accordingly slope estimates were associated with minimal bias and MSE across different levels of censoring and that no significant variation in these values was noticed across the different levels of probabilities. Our simulation results also suggested that introducing covariates into the model did not increase the bias and MSE significantly compared to the model that just adjusts for informative right censoring without any covariate [2]. Secondly is the effect of the value of probability of success π on the model's performance in terms of bias and accuracy. As noticed from Table 1, when $\pi = 0$ the overall bias was -0.076, with $\pi = 0.25$ the overall bias increased to -0.079, with $\pi = 0.5$ the overall bias decreased to about -0.076 and with $\pi = 0.75$ bias decreased to -0.071. As for the accuracy of the estimated population slope corresponding to the covariate which is assessed by $MSE(a)$, we realized that with $\pi = 0$ the overall $MSE(a)$ was about 0.017, with $\pi = 0.25$ it increased to about 0.019, whereas with $\pi = 0.5$, the overall $MSE(a)$ decreased to 0.015 and

with $\pi = 0.75$ it was about 0.16. As for $\text{MSE}(b)$, its overall value with $\pi = 0$ was 0.011, with $\pi = 0.25$ it was about 0.011, and 0.013 with $\pi = 0.5$ and with $\pi = 0.75$ it was about 0.014. Accordingly no significant association was denoted for the probability of success on the model's performance. Thus the probability of being in one group versus the other did not have any effect on the accuracy of the estimates. The accuracy of the model was also assessed using normally distributed covariates across different levels of censoring. Our results showed that at different levels of censoring bias, MSE_a and MSE_b were minimal with average values of 0.081, 0.0187, and 0.0152 respectively.

In summary our simulation results suggest that there wasn't any significant change in the accuracy and validity of the estimates when covariates were introduced to the model that adjusts for informative right censoring. Moreover, probability of success $\Pr(X_i = 1) = \pi$ i.e. probability of being in one group versus the other did not impact the accuracy of the estimates, neither did the level of censoring.

6. Renal Transplant Dataset

The discrete geometric model was applied on the cohort of renal transplantation and estimates of the slopes for covariates of interest adjusting for informative right censoring were obtained. Blood urea nitrogen (BUN) a biomarker used to assess renal function was the outcome of interest in this study. If BUN levels are decreasing with time then this could be an indication that the patient's condition is improving.

The impact of covariates of interest on the rate of change of BUN levels over time was assessed. In particular, interest was in assessing the rate of change for BUN levels for female patients compared to males and those for Caucasian patients compared to African-Americans. Furthermore, interest was also in estimating the rate of change of one age group versus another to assess whether the patient's age has any effect on the success rate of the transplant. In addition, the effect of donors' characteristics on the renal function of the recipients following the transplant was also investigated. For this purpose, we estimated the rate of change of BUN levels for the patients whose donors were from the same race and same gender, in addition to assessing the impact of the donors vital status. These estimates are of significance since they can help clinicians decide on the best candidate to receive the graft after evaluating the characteristics of both donor and recipient.

In this study, patients who had undergone renal transplant in the year 2000 at the Medical University of South Carolina were identified. Demographical information and BUN levels were collected on 109 patients who had kidney transplant in calendar year 2000. Seven repeated measurements on BUN were recorded at 0, 1, 3, 6, 12, 24, and 36 months following the transplant; i.e., three-year follow-up measurements were collected on BUN between the years 2000 to 2003. Detailed description of this dataset can be found in [3]. Normal BUN levels are usually in the range of 0.7-1.3 mg per 100 ml of blood. About 20% of the observations had informative right censoring due to graft failure. A logarithmic transformation was applied on BUN and the follow-up times to achieve linearity. The mean BUN levels stratified by the number of observations were plotted against the follow-up months on a logarithmic scale (Figure 1). As depicted in Figure 1, patients with fewer

number of follow-up visits have higher means of BUN levels compared to those with more follow-up visits. This indicates that patients who experienced graft failure have higher BUN levels and fewer number of follow-up visits. Population slope estimates of the BUN corresponding to the covariates of interest were generated using SAS procedure NLMIXED [13] and the results are shown in Table 2. Parameter estimates and their standard errors (SE) were computed using the second derivative matrix in the likelihood function. Accordingly t statistic is the ratio of the parameter estimate and its associated SE. P-values were then obtained using the t statistics with degrees of freedom equal to the number of subjects minus the number of random effects.

7. Results of the application on the renal transplant dataset

Our results showed that for female patients, the rate of change for BUN has decreased by 0.0517 (P-value = 0.0002) compared to males. The censoring parameter was estimated and its value was $\gamma_1=5.74$ (P-value < 0.0001) indicating that the censoring was informative. This result suggests that the renal function for the female patients is likely to improve with time, after the transplant, better than male patients. In our study we also assessed the effect of the patients race on the progression of the renal function. Our results showed that the BUN level for Caucasians has decreased by 0.063 (P-value < 0.0001) compared to African-Americans with estimated censoring of $\gamma_1=5.65$ (P-value < 0.0001). Thus, this result indicates that the censoring was informative and that following the renal transplant, the renal function for Caucasian patients performed better than that for African-Americans. As for the effect of the patient's age, our results also showed that the patient age has an effect on the renal function after the transplant. The rate of change for BUN corresponding to patients whose ages were under 35 was shown to decrease by 0.0646 (P-value = 0.001) compared to patients whose ages were 35 and over. The estimated censoring parameter is $\gamma_1 = 5.706$ (P-value < 0.0001). This result suggests that after the transplant, the kidneys for patients under 35 function better than those over age 35. Moreover, we investigated the effect of age as a continuous covariate on the rate of change for BUN and our results indicated that when a patient is older by one year BUN level is increased by 0.0047 mg per 100 ml of blood (P-value < 0.0001). Accordingly, our results suggest that female patients performed better than males, and that Caucasian patients did better than African-Americans. Furthermore, for patients under the age of 35, kidneys functioned better than those over 35 following transplantation.

In our study, we also tried to address an important question that is always raised by the clinicians as to what patient should the kidney be transplanted. In this context we tried to address the question of who is a better candidate to receive the kidney after taking into account the characteristics of the donor and the receiver. To answer this question we examined the characteristics of the donor in terms of assessing the effect of donor's gender, race, and vital status whether living or deceased on the rate of change of BUN following the renal transplant. More importantly, we examined whether having the same race between the donor and patient has any effect on the success of the transplantation. We also assessed the effect of having the same donor-recipient gender on the BUN levels. Our results showed that male donors are better than female donors. This is attributed to the fact that BUN levels corresponding to patients whose donors were males has decreased by 0.0565 (P-value < 0.0001) compared to patients whose donors were females. The corresponding censoring

parameter was estimated by $\gamma_1 = 6.19$ (P-value < 0.0001), indicating that the censoring was informative. This result suggests that male donors are better than female donors irrespective of the gender of the recipient. Our results also showed that if the donor was living then this has a better effect on the renal function compared to when the donor was deceased. The BUN levels for patients whose donors are living have decreased by 0.074 (P-value < 0.0001) compared to patients whose donors were deceased. The censoring parameter was estimated by $\gamma_1 = 5.81$ (P-value < 0.0001) indicating that the censoring process was informative. As for donors' race, our results showed that patients whose donors were Caucasians did better than other patients whose donors were from a different race with estimated BUN slope of -0.035 . However, this result was not statistically significant (P-value = 0.08). All the results discussed thus far indicate that the donor characteristics have an effect on the renal outcome for the patients after the transplantation. In this study we also investigated whether having the same donor-recipient gender has any effect on the renal function. Our results showed that if the donor was male and the recipient was also male, then BUN levels decrease by 0.073 (P-value = 0.0004) with $\gamma_1 = 4.54$ (P-value < 0.0001) compared to male donors and female recipients. On the other hand, when donors and recipients were both females, the rate of change of BUN decreased by 0.0059 compared to the case of female donors and male. However this result was statistically insignificant with P-value of 0.84. Accordingly, for male donors better kidney outcomes were reached for male recipients than females. In this study we also investigated if having the same race between donors and recipients had any effect on the renal function after the transplantation. Our results suggest that if the donor and the recipient had the same race then the BUN level decreases by 0.16 (P-value < 0.0001) with $\gamma_1 = 5.7$ (P-value < 0.0001) compared to the case when the donor and the recipient had different races. In this context we investigated the case when donors were Caucasians and the recipients were also Caucasians versus the case when donors were Caucasians but the recipients were African-Americans. Our results showed that when both donors and recipients were Caucasians, then the kidneys performed better than when donors were Caucasians and recipients were African-Americans. The associated rate of change for BUN has decreased by 0.21 (P-value < 0.0001) and censoring parameter was estimated by $\gamma_1 = 4.91$ (P-value < 0.0001). Thus our results suggest that when the donor is Caucasian then if a Caucasian patient was assigned to receive the transplant then this patient is likely to have a greater decrease in BUN levels after the transplantation compared to African-American recipients. We also investigated the case when the donors were African-Americans and the recipients were African-Americans. Our results showed that even in the case when the donors were African-Americans, the African-American patients did not do well after the transplantation since their associated BUN level did not decrease after the transplantation. On the contrary, their associated BUN levels has significantly increased by 0.38 compared to Caucasian receivers (P-value < 0.0001 and $\gamma_1 = 10.72$; P-value < 0.0001).

The mean of the predicted individual subject slopes β_i 's for all the covariates were plotted against each number of observations m_i (Figure 2). This Figure shows that individuals who had number of observations of two, three or four did not on average experience a decrease in their BUN levels over the three-year period following transplantation. Whereas, individuals who had greater number of observations appear to experience on average a decrease in their

BUN levels. This confirms, in addition to having a significant censoring parameter γ_1 , that the censoring process is dependent on the individual subject slopes for BUN levels.

8. Conclusion

In this study, we developed a new model that estimates slopes for a longitudinally recorded outcome adjusting for informative right censoring. The estimated slope adjusts for the covariates effects. This model is very important since it enables the clinicians and researchers to assess the progression of a certain disease adjusted for a particular covariate of interest. Moreover, the proposed model based on the simulation results showed to have accurate estimates compared to the model that does not accommodate the covariate effect and adjusts solely for the informative right censoring [2]. Our model was applied to the renal transplant patients and our results suggested interesting findings. The results of our study indicated that for patients who have undergone renal transplant various factors could affect the success of this surgery. Among these factors are gender, race, donor's vital status and the match between donors and recipients. These results are in agreement with our previous findings [3]. However, the approach we adopted in the current manuscript enabled us to overcome the convergence problem encountered in our pervious study [3] that was due to the small sample size resulting from stratification on the different covariates. This was achieved by incorporating the covariates in likelihood function and implementing direct maximization of the marginal likelihood function.

Acknowledgments

This work was supported by National Institutes of Health Grants HL077192 and HL087986

References

1. Fletcher, R. Practical Methods of Optimization. 2nd. Wiley; New York: 1987.
2. Jaffa MA, Woolson RF, Lipsitz SR. Slope estimation for bivariate longitudinal outcomes adjusting for informative right censoring by using a discrete survival model: application to the renal transplant cohort. *J Royal Statistical Society: Series A (Statistics in Society)*. 2011; 174:387–402.
3. Jaffa MA, Woolson RF, Lipsitz SR, Baliga PK, Lopes-Virella M, Lackland DT. Analyses of renal outcome following transplantation adjusting for informative right censoring and demographic factors: A longitudinal study. *Renal Failure*. 2010; 32:691–698. [PubMed: 20540637]
4. Jennrich RI, Schluchter MD. Unbalanced repeated measures models with structured covariance matrices. *Biometrics*. 1986; 42:805–820. [PubMed: 3814725]
5. Laird NM. Missing data in longitudinal studies. *Stat Med*. 1988; 44:175–188.
6. Laird NM, Wang F. Estimating rates of change in randomized clinical trials. *Controlled Clinical Trials*. 1990; 11:405–419. [PubMed: 1963133]
7. Laird NM, Ware JH. Random-effects models for longitudinal data. *Biometrics*. 1982; 38:963–974. [PubMed: 7168798]
8. Linstrom MJ, Bates MD. Newton-Raphson and EM algorithms for linear mixed-effects models for repeated measures data. *J Amer Stat Asso*. 1988; 83:1014–1022.
9. Mori M, Woolson RF, Woodworth GG. Slope estimation in the presence of informative right censoring: modeling the number of observations as a geometric random variable. *Biometrics*. 1994; 50:39–50. [PubMed: 8086614]
10. Pinheiro JC, Bates DM. Approximations to the log-likelihood function in the nonlinear mixed - effects model. *Journal of Computational and Graphical Statistics*. 1995; 4:12–35.

11. Reed E, Cohen CJ, Barr ML, Ho E, Reemtsma K, Rose EA, Hardy M, Suciu-Foca N. Effect of recipient gender and race on heart and kidney allograft Survival. *Transplant Proc.* 1992; 24:2670–2671. [PubMed: 1465895]
12. Richie RE, Niblack GD, Johnson HK, Green WF, MacDonnell RC, Turner BI, Tallent M. Factors influencing the outcome of kidney transplants. *Ann Surg.* 1983; 19:672–677. [PubMed: 6344816]
13. SAS Institute Inc.. *SAS/STAT® 9.2 User's guide.* Cary, NC, USA: 2002.
14. Schluchter MD. Methods for the analysis of informatively censored longitudinal data. *Stat Med.* 1992; 11:1861–1870. [PubMed: 1480878]
15. Stefanski LA, Carroll RJ. Covariate measurement error in logistic regression. *Annals of Statistics.* 1985; 13:1335–1351.
16. Vonesh EF, Carter RL. Efficient estimates for random-coefficient growth curve models with unbalanced data. *Biometrics.* 1987; 43:617–628. [PubMed: 3663819]
17. Wu MC. Sample size for comparison of changes in the presence of right censoring caused by death, withdrawal, and staggered entry. *Controlled Clinical Trials.* 1988; 9:32–46. [PubMed: 3356151]
18. Wu MC, Bailey K. Estimation and comparison of changes in the presence of informative right censoring: conditional linear model. *Biometrics.* 1989; 45:939–955. [PubMed: 2486189]
19. Wu MC, Carroll RJ. Estimation and comparison of changes in the presence of informative right censoring by modeling the censoring process. *Biometrics.* 1988; 44:175–188.

Figure 1.

Mean natural logarithm of blood urea nitrogen against natural logarithm of follow-up months following transplantation stratified by number of observations. Figure legend show number of observations.

Figure 2.
Mean of the predicted individual slopes Beta_i for all the covariates plotted against each number of observations m_i .

Table 1

Summary * of Performance of the likelihood estimator assuming number of observations follows a Geometric model and covariate X_i follows I) a Bernoulli distribution with probability of success π equals: 1) 0; 2) 0.25; 3) 0.5; 4) 0.75 and II) a normal distribution with mean 43.51 and standard deviation of 13.9.

Censoring Parameter	Performance indicator	I) Probability of success				II) Normal
		1) $\pi = 0$	2) $\pi = 0.25$	3) $\pi = 0.5$	4) $\pi = 0.75$	$\mu=43.51, \sigma=13.9$
$\gamma_1 = 0$	Bias	-0.075	-0.077	-0.073	-0.067	0.081
	MSE(a)	0.017	0.024	0.016	0.017	0.020
	MSE(b)	0.011	0.012	0.014	0.015	0.016
$\gamma_1 = 2.3$	Bias	-0.078	-0.077	-0.074	-0.069	0.082
	MSE(a)	0.016	0.018	0.016	0.017	0.019
	MSE(b)	0.011	0.012	0.013	0.015	0.016
$\gamma_1 = 4.5$	Bias	-0.077	-0.080	-0.080	-0.074	0.081
	MSE(a)	0.019	0.017	0.014	0.015	0.018
	MSE(b)	0.011	0.012	0.013	0.014	0.015
$\gamma_1 = 6.2$	Bias	-0.077	-0.083	-0.078	-0.074	0.083
	MSE(a)	0.017	0.015	0.015	0.015	0.018
	MSE(b)	0.010	0.011	0.012	0.014	0.014

* Outcomes were simulated from a normal distribution with parameters $\theta_0 = -0.044$, $\theta_1 = -0.052$, $\sigma\beta=0.065$, and $\sigma_\varepsilon=0.35$. Number of observations were simulated from the truncated geometric distribution with parameters $\gamma_0 = -3.46$. Number of simulations was 2000.

Table 2

Estimated population slope* and censoring parameter** corresponding to the covariates: 1) patient's gender, 2) patient's race, 3) patient's age, 4) patients age in years 5) donor's race, 6) donor's gender, 7) donor's vital status (alive/deceased), 8) same donor-patient race vs. different donor-patient race, 9) male donor-patient vs. male donor-female patient 10) female donor-patient vs. female donor-male patient, 11) Caucasian donors-Caucasian patients vs. Caucasian donors-African-American patients, and 12) African-American Donors and patients vs. African-American donors-Caucasian patients.

Recipients/donors characteristics	N	θ_i^*	SE	P-value	χ^{**}	P-value
1) Female Recipients	46	-0.052	0.0134	0.0002	5.74	< 0.0001
2) Caucasian Recipients	47	-0.063	0.0156	< 0.0001	5.65	< 0.0001
3) Recipients < 35	32	-0.065	0.0189	0.001	5.7	< 0.0001
4) Recipients age	109	0.0047	0.0006	< 0.0001	5.7	< 0.0001
5) Caucasian Donors	80	-0.035	0.0189	0.089	5.56	< 0.0001
6) Male Donors	72	-0.057	0.0139	< 0.0001	6.2	< 0.0001
7) Alive Donors	21	-0.074	0.0187	0.0001	5.81	< 0.0001
8) Same Race Donors- Recipients	62	-0.16	0.0221	< 0.0001	5.72	< 0.0001
9) Male Donors-Male Recipients	41	-0.073	0.0195	0.0004	4.54	< 0.0001
10) Female Donors-Female Recipients	15	-0.0059	0.0293	0.84	12.84	< 0.0001
11) Caucasian Donors-Caucasian Recipients	42	-0.21	0.0298	< 0.0001	4.91	< 0.0001
12) African-American Donors and Recipients	19	0.375	0.039	< 0.0001	10.72	< 0.0001