

Research Article

Two-Stage Joint Model for Multivariate Longitudinal and Multistate Processes, with Application to Renal Transplantation Data

Behnaz Alafchi ¹, Hossein Mahjub ², Leili Tapak ³, Ghodratollah Roshanaei ³,
and Mohammad Ali Amirzargar ⁴

¹Department of Biostatistics, School of Public Health, Hamadan University of Medical Sciences, Hamadan, Iran

²Research Center for Health Sciences, Department of Biostatistics, Faculty of Public Health, Hamadan University of Medical Sciences, Hamadan, Iran

³Department of Biostatistics, School of Public Health, Modeling of Noncommunicable Diseases Research Center, Hamadan University of Medical Sciences, Hamadan, Iran

⁴Department of Urology, Ekbatan Medical Center, Hamadan University of Medical Sciences, Hamadan, Iran

Correspondence should be addressed to Ghodratollah Roshanaei; gh.roshanaei@umsha.ac.ir

Received 21 October 2020; Revised 16 March 2021; Accepted 23 March 2021; Published 9 April 2021

Academic Editor: Hyungjun Cho

Copyright © 2021 Behnaz Alafchi et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In longitudinal studies, clinicians usually collect longitudinal biomarkers' measurements over time until an event such as recovery, disease relapse, or death occurs. Joint modeling approaches are increasingly used to study the association between one longitudinal and one survival outcome. However, in practice, a patient may experience multiple disease progression events successively. So instead of modeling of a single event, progression of the disease as a multistate process should be modeled. On the other hand, in such studies, multivariate longitudinal outcomes may be collected and their association with the survival process is of interest. In the present study, we applied a joint model of various longitudinal biomarkers and transitions between different health statuses in patients who underwent renal transplantation. The full joint likelihood approaches are faced with the complexities in computation of the likelihood. So, here, we have proposed two-stage modeling of multivariate longitudinal outcomes and multistate conditions to avoid these complexities. The proposed model showed reliable results compared to the joint model in case of joint modeling of univariate longitudinal biomarker and the multistate process.

1. Introduction

In longitudinal medical studies, repeated measurements of biomarkers are usually collected until the occurrence of a clinical event such as recovery, disease relapse, or death. The main objective of these studies is to study the association between two correlated processes or to use the information of biomarkers to predict or explain the time-to-event. In such analyses, joint models are needed to accurately study the association between two processes. Joint modeling approaches, which have been developed to handle the association between a single

longitudinal biomarker and a time-to-event outcome, have received considerable attention recently [1–3].

In practice, a patient may experience multiple disease progression events prior to the event of interest; for example, the patient with cancer may experience a local recurrence followed by a distant recurrence and then death. So, instead of the occurrence of a single event, the progression of the disease should be modeled as a multistate process, focusing on the transitions between different health states and the impact of the longitudinal biomarker on them. Moreover, in such studies, multiple longitudinal biomarkers may be

collected for each patient and the correlation structure between them should be taken into account in the model [4].

There are several studies that have extended classical joint models for when there are multiple longitudinal outcomes. Chi and Ibrahim [5] extended a joint model for multivariate longitudinal data and multivariate survival data. Then, Andrinopoulou et al. [6] extended a joint model for competing risks and two longitudinal biomarkers. On the other hand, there are few studies that extended the joint models for multistate data. Dantan et al. [7] proposed a joint multistate model with the latent state for one longitudinal response and illness-death data. Ferrer et al. [4] developed a joint model with the shared random effect framework for joint modeling of longitudinal and multistate processes. However, based on our knowledge, there is no study to model multivariate longitudinal and multistate processes jointly. So, in this paper, we are interested to simultaneous modeling of these two correlated processes.

The joint model based on the shared random effects framework, which is the most commonly used framework of joint modeling of longitudinal and survival data, is faced with implementation issues when the number of the longitudinal outcomes grows [8]. Moreover, modeling the transitions between different health states instead of the time to a single event can make the issue more complicated.

The primary joint models have been based on the two-stage approaches in which the likelihood calculation is implemented in two steps [1, 8, 9]. These initial approaches are not faced with the problem of computational difficulties on the full joint likelihood calculation [1]. So, here, we use the idea of these initial approaches to avoid the computational complexities and propose a two-stage base model for joint modeling of multivariate longitudinal and multistate data. The rest of the paper is organized as follows. The motivating example is described in Section 2. The basic concepts of joint modeling of longitudinal and survival data are provided in Section 3. In Section 4, we illustrate our proposed two-stage based model for multivariate longitudinal and multistate data. The application of the model to renal transplantation data set is presented in Section 5, and a discussion and conclusion is finally given in Section 6.

2. Motivating Dataset

Our motivating example includes end-stage renal disease (ESRD) patients who underwent renal transplantation. The information of all renal transplantations performed at Shahid Beheshti hospital in Hamadan province (western of Iran) from July 1994 to February 2017 ($n=408$) was collected retrospectively. Patients were followed from the time of transplantation at regular visits (first week, first month, third month, sixth month, and every year thereafter). The number of visits for each patient was variable with the median number of 12 visits and varied between 3 and 37 visits. Donor-Recipient Gender (DRG) was matched in 221 (54.2%) transplantations and was mismatched in 187 (45.8%) transplantations. DRG match included both male donor to male recipient and female donor to female recipient, though DRG mismatch included female donor to

male recipient and vice versa. The mean (\pm SD) age of patients was 37.24 ± 13.46 years.

Several parameters, so-called biomarkers, such as creatinine, hemoglobin, and blood urea nitrogen (BUN), were collected in regular visits to monitor renal transplantation. These biomarkers are variably associated with renal transplantation function and mortality. Often, it is the trend of these biomarkers rather than the measure of their level at a specific time point that affect the risk of an event. Moreover, acute or chronic graft rejection is an unwanted outcome of transplantation, which can affect the overall survival of patients. Indeed, a patient may experience graft failure rejection prior to death and this graft failure may cause an increase in the risk of mortality. So, in this study, there were two types of outcomes, including (1) multiple longitudinal outcomes represent the longitudinal biomarkers and (2) multistate outcome that represents the survival process. To take into account the correlation between longitudinal biomarkers as well as the correlation between longitudinal and multistate processes, it was essential to use the joint modeling approaches.

Multistate process is shown in Figure 1. After transplantation (state 1), a patient may experience graft rejection (state 2) and death (state 3) due or not to ESRD. It should be noted that here all patients with acute or chronic rejection were considered as the same states, due to the small sample sizes. The Ω matrix represents the amount of direct observed transitions between different states. In total, the graft of 53 patients was rejected after transplantation. Overall, 49 patients have died during the follow-up; among them, 14 patients died after rejection, and 35 patients died without graft rejection. Moreover, 359 patients were censored during the follow-up; among them, 320 patients were censored for both rejection and death events, and 39 patients were censored for death after rejection. These 39 patients were returned to dialysis.

$$\Omega = \begin{matrix} & \begin{matrix} 1 & 2 & 3 \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \end{matrix} & \begin{pmatrix} 320 & 53 & 35 \\ 0 & 39 & 14 \\ 0 & 0 & 49 \end{pmatrix} \end{matrix}. \quad (1)$$

The observed levels of the longitudinal biomarkers are called observed values and the predicted levels of longitudinal biomarkers based on the multivariate linear mixed effects model are called expected values. The mean of observed and expected values of the longitudinal biomarkers over time are depicted in Figure 2. In this figure, patients are divided into four groups: those who experienced graft rejection, those whose graft was not rejected during the follow-up, those who died, and those who were alive at the end of the study.

3. Simultaneous Modeling of Longitudinal and Time-to-Event Data

In this section, we will briefly describe the joint modeling of a single longitudinal biomarker and time to a single event of

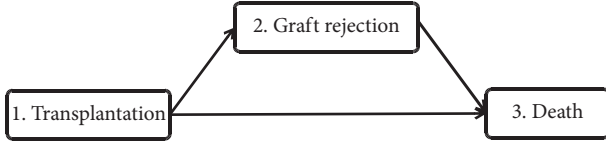


FIGURE 1: The structure of multistate data.

interest, in the shared random effects framework. This model consists of two submodels, including a linear mixed submodel for longitudinal data (repeated measurements of the longitudinal biomarker) and a proportional hazard submodel for time-to-event data.

3.1. Longitudinal Submodel. Let us assume that Y_{ij} is the measure of a biomarker on subject $i = 1, 2, \dots, n$, at time t_{ij} , $j = 1, 2, \dots, n_i$. The trajectory of this longitudinal variable is modeled by a linear mixed model as follows:

$$Y_{ij} = Y_i(t_{ij}) = X_i^L(t_{ij})^T \beta + Z_i(t_{ij})^T \nu_i + \varepsilon_{ij}, \quad (2)$$

where $X_i^L(t_{ij})$ and $Z_i(t_{ij})$ are the vector of covariates associated with the vector of fixed effects β and the vector of random effects ν_i , respectively, where T denotes the transpose operator. We assumed that $\nu_i = (\nu_{0i}, \nu_{1i}) \sim N(0, \Sigma)$ (ν_{0i} and ν_{1i} are the random intercept and random slope effects, respectively) and $\varepsilon_i = (\varepsilon_{i1}, \varepsilon_{i2}, \dots, \varepsilon_{in_i})^T \sim N(0, \sigma^2 I_{n_i})$, where I is the identity matrix of order n_i ; ν_i and ε_i are independent [10].

3.2. Proportional Hazard Submodel. Let $T_i = \min(T_i^*, C_i)$ represent the observed failure time for the i th subject, where T_i^* denotes the true failure time, and C_i is the censoring time for i th subject. Moreover, an event indicator is defined as $\delta_i = I(T_i^* \leq C_i)$, which equals 1 if the failure is observed and 0 otherwise. To model the time-to-event occurrence, a proportional hazard Cox model [11], which takes into account the biomarker's dynamics via shared random effects, ν_i , is used. Thus, the model is defined as follows:

$$\begin{aligned} \lambda_i(t|\nu_i) &= \lim_{dt \rightarrow 0} \frac{\Pr(t \leq T_i^* < t + dt | T_i^* > t, X_i^S, \nu_i)}{dt} \\ &= \lambda_0(t) \exp\{X_i^S(t)^T \gamma + \eta \omega_i(t, \nu_i)\}, \end{aligned} \quad (3)$$

where $\lambda_0(t)$ is the hazard at baseline and $X_i^S(t)$ is the vector of possibly time-dependent fixed effects covariates associated with the vector of coefficients γ . The multivariate function $\omega_i(t, \nu_i)$ defines the association between the longitudinal and survival processes, and η is the coefficient that quantifies this association. Here, different (most commonly used) association functions which are used to describe the association structure between longitudinal and time-to-event data are presented as follows (note, for ease of writing, here only the time variable is included in the model; however other covariates can also be included in the model):

- (i) The association structure takes into account the random time trend in the time-to-event model. This association structure is defined as

$$\omega_i(t, \nu_i) = \nu_{0i} + \nu_{1i}t. \quad (4)$$

So, the time-to-event submodel is rewritten as

$$\lambda_i(t|\nu_i) = \lambda_0(t) \exp\{X_i^S(t)^T \gamma + \eta(\nu_{0i} + \nu_{1i}t)\}, \quad (5)$$

in which η represents the association between the longitudinal biomarker measurements and the risk of the event at time t [2]. This association structure expresses the subject-specific deviations from the average intercept and average slope.

- (ii) The association structure includes the true value of the longitudinal biomarker (at time t) in the time-to-event model, as

$$\omega_i(t, \nu_i) = \beta_0 + \beta_1 t + \nu_{0i} + \nu_{1i}t. \quad (6)$$

So, the time-to-event submodel is rewritten as

$$\lambda_i(t|\nu_i) = \lambda_0(t) \exp\{X_i^S(t)^T \gamma + \eta(\beta_0 + \beta_1 t + \nu_{0i} + \nu_{1i}t)\}, \quad (7)$$

in which η is the association between the longitudinal biomarker and the risk of the event at time t taking into account fixed and random effects predictions of the true value of the longitudinal biomarker [1].

- (iii) The association structure includes both the true value of the longitudinal biomarker and the slope of the true biomarker's trajectory at time t in the time-to-event model. This association structure is defined as

$$\begin{aligned} \omega_i(t, \nu_i) &= \beta_0 + \beta_1 t + \nu_{0i} + \nu_{1i}t, \\ \omega_i'(t, \nu_i) &= \frac{d}{dt} (\beta_0 + \beta_1 t + \nu_{0i} + \nu_{1i}t). \end{aligned} \quad (8)$$

So, the time-to-event submodel is rewritten as

$$\begin{aligned} \lambda_i(t|\nu_i) &= \lambda_0(t) \exp\{X_i^S(t)^T \gamma + \eta_1(\beta_0 + \beta_1 t + \nu_{0i} + \nu_{1i}t) \\ &\quad + \eta_2(\beta_1 + \nu_{1i})\}, \end{aligned} \quad (9)$$

in which η_1 represents the association between the current true value of longitudinal biomarker and the risk of the event at time t like the previous association function, and η_2 defines the association between the slope of the true trajectory at time t and the risk of the event. Note that this association structure can catch situations in which two patients have a similar true level of the biomarker at time t , but they may have different rate of change of the biomarker [12].

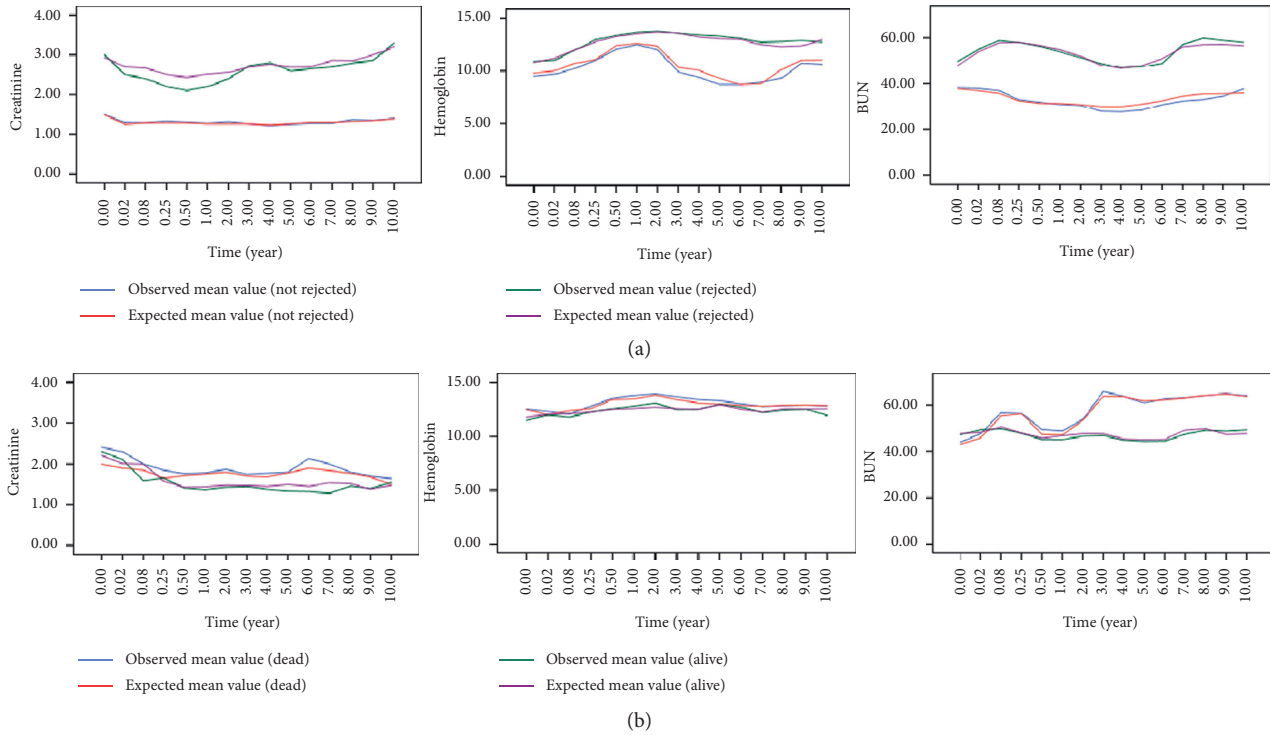


FIGURE 2: The mean of observed value of the longitudinal biomarkers (observed mean value) and the mean value of their predicted value based on the multivariate mixed effects model (expected mean value) over time based on (a) graft rejection status and (b) death state.

(iv) The association structure takes into account the cumulative effect of the longitudinal biomarker (the area under the biomarker trajectory) in the time-to-event model. In this association structure, the integral of the longitudinal trajectory that represents the cumulative effect of the biomarker up to time t is calculated via

$$\omega_i(t, \nu_i) = \int_0^t (\beta_0 + \beta_0 s + \nu_{0i} + \nu_{1i} s) ds. \quad (10)$$

So, the time-to-event submodel is rewritten as

$$\lambda_i(t|\nu_i) = \lambda_0(t) \exp \left\{ X_i^S(t)^T \gamma + \eta \left(\int_0^t (\beta_0 + \beta_0 s + \nu_{0i} + \nu_{1i} s) ds \right) \right\}, \quad (11)$$

in which η represents the association between the whole history of the longitudinal biomarker up to time t and the risk of the event at time t . Note that, in this association structure, the whole history of the biomarker is associated with the risk of the event, while in the three previous structures, the risk of an event is just allowed to be associated with the features of the biomarker at a specific time point [1].

4. Simultaneous Modeling of Multivariate Longitudinal and Multistate Data

As there are multiple longitudinal biomarkers with a multistate process in our motivating dataset, the clinical interest is on the correlation between the longitudinal biomarkers and their association with transitions between different health states. However, the implementation of the joint

model within the shared random effects framework, mentioned in the previous section, is difficult when the number of biomarkers is large. This computational problem is due to the high number of parameters in the variance-covariance matrix of the random effects. Moreover, modeling the transitions between different health states instead of the time to a single event can make the issue more complicated. In order to overcome this issue, a quick and approximate estimation method is required.

The initial joint models are based on the two-stage methods, where the likelihoods of the longitudinal and time-to-event data are calculated separately [12–14]. Some investigators have shown in some particular studies that the performance and the resulting estimations of these initial two-stage based models are similar to the joint models (based on the full joint likelihood calculation) [15] in the case of a single longitudinal outcome and single time-to-event

outcome. So, in the current paper, we will propose an extension of the joint models for multivariate longitudinal and multistate data based on the main idea of the two-stage approaches.

The proposed model is implemented in two stages. At the first stage, a multivariate linear mixed model is used for multiple longitudinal biomarkers. Then, at the second stage, a multistate model with proportional hazards that incorporate the dynamics of each longitudinal biomarker through association structures (Section 3.2) is used for multistate data.

4.1. Stage 1: Multivariate Linear Mixed Effects Model. At the first stage, a multivariate linear mixed effects model is used for modeling the longitudinal biomarkers measurements. The multivariate linear mixed effects model is defined as

$$Y_{ij,g} = Y_{i,g}(t_{ij,g}) = X_{i,g}^L(t_{ij,g})^T \beta_g + Z_{i,g}(t_{ij,g})^T \nu_{i,g} + \varepsilon_{ij,g}, \quad (12)$$

where $X_{i,g}^L(t_{ij,g})$ and $Z_{i,g}(t_{ij,g})$ are the vector of covariates associated with the vector of fixed effects β_g and the vector of random effects $\nu_{i,g}$ for the g th biomarkers, respectively. It is assumed that $\text{vec}(\nu_i) \sim N(0, \psi)$ when $\nu_i = (\nu_{i,1}, \nu_{i,2}, \dots, \nu_{i,l})$, $\nu_{i,l} = (\nu_{0i,l}, \nu_{1i,l})$ and $l = 1, 2, \dots, g$ (the ‘‘vec’’ operator vectorizes a matrix by stacking its columns). Note that ε_i is a $n_i \times g$ matrix in which the rows are independently distributed as $N(0, D)$, where n_i is the number of longitudinal measurements for subject i . To estimate fixed and random effects coefficients, maximum likelihood and empirical Bayes methods are used, respectively [11]. The main interest of this stage is to obtain the prediction of the longitudinal biomarkers from the multivariate mixed effects model to be included as covariates in the multistate model at Stage 2.

4.2. Stage 2: Multistate Model. At this stage, a multistate proportional hazards model, including the estimation of the longitudinal biomarkers from Stage 1, is fitted to the data. In the following, we describe the multistate process and Markovian multistate model which is used at this stage.

For each individual i , a multistate process is observed. Let $E_i = \{E_i(t), t \geq 0\}$ be the multistate process that takes values in finite state space $S = \{0, 1, \dots, M\}$ and denotes the state which is occupied by subject i at time t . The multistate process is assumed to be continuous, and E_i is a non-homogeneous Markov process. The Markov assumption ensures that the future of the process depends only on the past via the current

state. Let define $T_i = (T_{i1}, \dots, T_{im_i})$ as the vector of the $m_i \geq 1$ observed transition times for the i th subject, with $T_{ir} < T_{i(r+1)}, \forall r \in \{0, \dots, m_i - 1\}$; i.e., the time of each transition should be greater than the previous one. The vector of observed transition indicators for subject i is defined by $\delta_i = (\delta_{i1}, \dots, \delta_{im_i})$, with δ_{ir} equal to 1 if a direct transition to state r is observed at time T_{ir} and 0 otherwise. There will be m_i direct transition for subject i , if the last observed state is an absorbing state (i.e., it is impossible to leave once it is entered (e.g., death state)). Otherwise, there are $m_i - 1$ direct transitions, and T_{im_i} is equal to the right censoring C_i .

A Markov multistate model with proportional hazards is used to model the transition times. The transition intensity for the transition from state $h \in S$ to state $k \in S$ at time t is given by

$$\lambda_{hk,i}(t|\nu_i) = \lambda_{hk,0}(t) \exp \left\{ X_{hk,i}^S(t)^T \gamma_{hk} + \sum_g \eta_{hk,g} \omega_{hk,i,g}(t, \nu_{i,g}) \right\}, \quad (13)$$

where $\lambda_{hk,0}(\cdot)$ is the completely unspecified baseline hazard for the transition from h to k , $X_{hk,i}^S(t)$ is the vector of possibly time-dependent transition-specific covariates, and γ_{hk} is the associated vector of fixed regression coefficients. The multivariate function $\omega_{hk,i,g}(t, \nu_{i,g})$ (mentioned in Section 3) defines the association between the g -th longitudinal biomarker and the transition from state h to k , and its corresponding coefficient $\eta_{hk,g}$ quantifies this association [4]. The maximum likelihood method is used to estimate the parameters of this model.

The details of the implementation procedure are given in Section 5, when the model is used to analyze the renal transplantation data.

5. Application to Renal Transplantation Data

In this section, we first describe the implementation of our proposed two-stage based model and then give its fitting results for our motivating renal transplantation data.

5.1. Implementation of the Two-Stage Model. Stage 1. A multivariate linear mixed effects model is fitted. Use the `mlmmm.em()` function from the `R mlmmm` package to fit the multivariate linear mixed effects model. Joining the following equations showed the multivariate mixed model for subject i at time j :

$$\begin{aligned} \text{Creatinine}_{ij} &= \beta_{0,1} + \beta_{1,1}t_{ij,1} + \beta_{2,1}t_{ij,1}^2 + \beta_{3,1}\text{DRG} + \beta_{4,1}\text{age} + \nu_{0i,1} + \nu_{1i,1}t_{ij,1} + \varepsilon_{ij,1}, \\ \text{Hemoglobin}_{ij} &= \beta_{0,2} + \beta_{1,2}t_{ij,2} + \beta_{2,2}t_{ij,2}^2 + \beta_{3,2}\text{DRG} + \beta_{4,2}\text{age} + \nu_{0i,2} + \nu_{1i,2}t_{ij,2} + \varepsilon_{ij,1}, \\ \text{BUN}_{ij} &= \beta_{0,3} + \beta_{1,3}t_{ij,3} + \beta_{2,3}t_{ij,3}^2 + \beta_{3,3}\text{DRG} + \beta_{4,3}\text{age} + \nu_{0i,3} + \nu_{1i,3}t_{ij,3} + \varepsilon_{ij,3}. \end{aligned} \quad (14)$$

As described in Section 3, different association structures can be used to explain the dependence of the two processes.

Here, we focus on the four association structures described in Section 3.2. So, the estimation of the parameters is used to

obtain the linear predictors of the longitudinal biomarkers via the following equations:

$$\begin{aligned}
 \omega_{hk,i,g}^1(t, \nu_{i,g}) &= \nu_{0i,g} + \nu_{1i,g}t, \\
 \omega_{hk,i,g}^2(t, \nu_{i,g}) &= \beta_{0,g} + \beta_{1,g}t + \beta_{2,g}t^2 + \beta_{3,g}\text{DRG} + \beta_{4,g}\text{Age} + \nu_{0i,g} + \nu_{1i,g}t, \\
 \omega_{hk,i,g}^3(t, \nu_{i,g}) &= \beta_{0,g} + \beta_{1,g}t + \beta_{2,g}t^2 + \beta_{3,g}\text{DRG} + \beta_{4,g}\text{Age} + \nu_{0i,g} + \nu_{1i,g}t, \\
 \omega_{hk,i,g}^3(t, \nu_{i,g}) &= (\beta_{1,g} + \beta_{2,g}t + \nu_{1i,g}), \\
 \omega_{hk,i,g}^4(t, \nu_{i,g}) &= \int_0^t (\beta_{0,g} + \beta_{1,g}s + \beta_{2,g}s^2 + \beta_{3,g}\text{DRG} + \beta_{4,g}\text{Age} + \nu_{0i,g} + \nu_{1i,g}s) ds.
 \end{aligned} \tag{15}$$

These linear predictors are used as the covariates in Stage 2 for the multistate model.

Stage 2. The transitions intensity between different states (shown in Figure 1) including the estimation of the

longitudinal biomarkers (linear predictors obtained at Stage 1) is defined as follows:

$$\lambda_{hk,i}(t|\nu_i) = \lambda_{hk,0}(t) \exp \left\{ \gamma_{1,hk}\text{DRG} + \gamma_{2,hk}\text{age} + \sum_{g=1}^3 \eta_{hk,g} \omega_{hk,i,g}(t, \nu_{i,g}) \right\}. \tag{16}$$

Note that, in the two-stage based models, the likelihood calculation for longitudinal and time-to-event submodel are done separately.

At this stage, the *mstate* package is used to fit the Markov multistate model. The *msprep()* and *expands.covs()* functions are used to prepare the multistate data, and then the *coxph()* function is used to estimate the parameters of the transition-specific multistate model.

5.2. Results. The results of the multivariate longitudinal model (Stage 1) are presented in Table 1. As shown in this table, the time variable had significant linear and quadratic effects on creatinine, hemoglobin, and BUN measurements. Furthermore, the results showed that DRG was associated with the creatinine trajectory and age of recipient was associated with hemoglobin, and BUN measurements (Table 1). Since the association between three biomarkers was at primary interest, the correlation of the random intercepts and random slopes of longitudinal biomarkers was examined in the multivariate model. The correlation matrix of the random effects based on joint modeling of the longitudinal biomarkers is given in Table 2. As shown in this table, the creatinine and BUN measurements were correlated on both the intercept ($r=0.67$) and the slope ($r=0.75$). Moreover, there was a moderate inverse correlation between the random slope of hemoglobin and the random slope of creatinine ($r=-0.52$) as well as between the random slope of hemoglobin and the random slope of BUN ($r=-0.55$).

The results of the multistate model (Stage 2) fitted with various association structures are given in Table 3. We can observe the effect of different association structures on the

results of two-stage based model through comparing their estimations. For instance, we had not found any significant effect of the true slope of creatinine and hemoglobin on time to graft rejection at time T_{12} (time of transition from state 1 to state 2), when we used their prediction based on the multivariate longitudinal model. However, the cumulative effect of these longitudinal biomarkers, as well as the effect of their true values on the mentioned transition, was significant. Such results display that the choice of association structure (mentioned in Section 2), which is used in the multistate model, is of particular importance and may lead to conflicting results.

As, in practice, the association structure is unknown, we can compare different association structures and select the best one via various model selection criteria. The log-likelihood values of the multistate models can be used to select the best structure. The values of the log-likelihood for different structures (presented in Table 3) revealed that including the estimation of the cumulative effect of longitudinal biomarkers in the multistate model had the best fit to data. It should be noted that choosing the appropriate association structure between longitudinal and multistate model also depends on the clinical aspects. Moreover, we can use different association structures for each of the longitudinal outcomes in a single multistate model and even in each of the transitions of a unique model.

Based on the results of the presented innovative method, we found that creatinine, hemoglobin, and BUN had predictive power in ESRD patients who underwent renal transplantation. The results showed that, after adjustments for fixed covariates, the cumulative effects (the whole area under the curve of the longitudinal biomarkers trajectories)

TABLE 1: The results of the multivariate longitudinal model (Stage 1).

Variables	Creatinine			Hemoglobin			BUN		
	β	SE	p -value	β	SE	p -value	β	SE	p -value
Intercept	1.689	0.129	<0.001	10.788	0.313	<0.001	26.932	2.946	<0.001
Time	-0.164	0.020	<0.001	1.048	0.039	<0.001	-3.159	0.355	<0.001
Time ²	0.016	0.007	0.022	-0.099	0.004	<0.001	0.235	0.093	0.011
DRG (matched)	0.085	0.042	0.043	-0.053	0.144	0.711	2.093	1.008	0.038
Recipient age	-0.001	0.002	0.517	0.016	0.005	0.004	0.156	0.050	0.002

TABLE 2: The correlation matrix of random effects for the random intercepts and random slopes of the multivariate joint model.

		Creatinine		Hemoglobin		BUN	
		Intercept	Slope	Intercept	Slope	Intercept	Slope
Creatinine	Intercept	1	-0.16	0.09	-0.40	0.67	-0.19
	Slope		1	-0.20	-0.52	-0.27	0.75
Hemoglobin	Intercept			1	-0.21	0.08	-0.16
	Slope				1	0.10	-0.55
BUN	Intercept					1	-0.70
	Slope						1

of creatinine (HR = 1.10; 95% CI: 1.01–1.18 and; $p = 0.024$), hemoglobin (HR = 0.90; 95% CI: 0.84–0.98 and; $p = 0.012$), and BUN (HR = 1.13; 95% CI: 1.02–1.24 and; $p = 0.016$) rather than their measurements at a particular time affected the risk of graft rejection. Moreover, the results revealed that the cumulative effect of hemoglobin on the survival after rejection was significant (HR = 1.13; 95% CI: 1.02–1.24 and; $p = 0.023$). The results also showed that an increase in age at the time of transplantation was associated with a lower risk of graft rejection (HR = 0.95; 95% CI: 0.91–0.99 and; $p = 0.013$), and a higher risk of death after rejection (HR = 1.14; 95% CI: 1.03–1.26 and; $p = 0.009$).

6. Discussion and Conclusion

Traditional joint models for longitudinal and time-to-event data often consist of one longitudinal and one survival outcome [1–3]. However, in longitudinal health studies, a patient may experience a succession of clinical events instead of a single event and multiple longitudinal biomarkers may be measured over time. Here, we proposed a two-stage based model for joint modeling of multivariate longitudinal and multistate data. With the use of two-stage based modeling, we avoid the computational complexities of likelihood calculation in the full joint likelihood approaches [8, 16]. Furthermore, as we have used multivariate longitudinal model that incorporated the correlation between biomarkers at the first stage and a multistate model for the disease progression process at the second stage, we have introduced an extension of the nave two-stage method in which univariate models are used at both stages.

The proposed model enabled us to study the complex association structure between multiple longitudinal biomarkers and transition between different health states in our

motivating renal transplantation dataset. The strategy of this proposed model was based on the specification of two separate submodels for multivariate longitudinal and multistate data. At the first stage, a multivariate linear mixed effects model proposed by Yücel [17] was used to model biomarkers trajectories. This model could handle the missing values which is a crucial issue in longitudinal data. Then, at the second stage, a multistate model with proportional hazards was used to model transition intensities incorporating the estimation of the random effects. In our case study, we assumed a Markov continuous multistate process as this assumption held for our renal data. However, a semi-Markov model, in which the hazard of entry into the next state depends on the sojourn time on the current state [18] or a non-Markov model [19], can be used simply instead. Moreover, as we did not have any information about graft retransplantation among patients who experienced graft rejection, we did not use a reversible multistate model, although this type of multistate model also can be used. So, the main advantage of the two-stage based proposed model is that it can be easily generalized to other types of longitudinal and survival models.

Guler et al. [8] have mentioned that the main limitation of two-stage methods is bias arising with the unignorable informative censoring in the longitudinal process which is caused by the survival process. Here, we compare joint and two-stage based models to see whether this bias is considerable. To achieve this aim, joint models for multistate and univariate longitudinal models (for each biomarker) are fitted and the results compare with two-stage based model. As presented in Appendix A (TableA.1), two approaches had similar likelihood values and the biases were minimal. Moreover, we showed that the impact of longitudinal biomarkers and the effects of covariates on the transition

TABLE 3: The results of the multistate models (Stage 2) for different association structures between longitudinal and multistate processes.

Association structure: true slope of longitudinal biomarkers ($\omega_{hk,ig}^1(t, \nu_{ig})$)									
Variables	Transition 1 to 2			Transition 2 to 3			Transition 1 to 3		
	Coef (SE)	HR (95% CI)	<i>p</i> -value	Coef (SE)	HR (95% CI)	<i>p</i> -value	Coef (SE)	HR (95% CI)	<i>p</i> -value
DRG	0.12 (0.11)	1.13 (0.91, 1.40)	0.258	0.22 (0.12)	1.24 (0.98, 1.57)	0.069	-0.32 (0.35)	0.73 (0.98, 1.57)	0.370
Recipient age	-0.01 (0.01)	0.99 (0.97, 1.01)	0.317	0.04 (0.02)	1.04 (1.00, 1.08)	0.045	0.005 (0.01)	1.00 (0.98, 1.03)	0.709
Creatinine	0.22 (0.12)	1.25 (0.98, 1.58)	0.067	0.49 (0.26)	1.63 (0.98, 2.72)	0.059	1.05 (0.54)	2.87 (0.99, 8.31)	0.052
Hemoglobin	-0.15 (0.08)	0.86 (0.74, 1.01)	0.061	-0.17 (0.09)	0.84 (0.71, 1.01)	0.058	0.36 (0.26)	1.44 (0.86, 2.42)	0.168
BUN	0.02 (0.01)	1.02 (1.00, 1.04)	0.045	-0.01 (0.01)	0.99 (0.98, 1.00)	0.143	0.01 (0.01)	1.01 (0.99, 1.03)	0.176
-2(loglikelihood)	5714.38								
Association structure: true value of longitudinal biomarkers ($\omega_{hk,ig}^2(t, \nu_{ig})$)									
Variables	Transition 1 to 2			Transition 2 to 3			Transition 1 to 3		
	Coef (SE)	HR (95% CI)	<i>p</i> -value	Coef (SE)	HR (95% CI)	<i>p</i> -value	Coef (SE)	HR (95% CI)	<i>p</i> -value
DRG	0.02 (0.11)	1.02 (0.83, 1.27)	0.837	-0.07 (0.12)	0.93 (0.74, 1.18)	0.554	-0.17 (0.37)	0.85 (0.41, 0.74)	0.651
Recipient age	-0.02 (0.02)	0.98 (0.94, 1.02)	0.317	0.12 (0.05)	1.13 (1.02, 1.24)	0.016	0.01 (0.02)	1.01 (0.97, 1.05)	0.617
Creatinine	0.45 (0.22)	1.57 (1.02, 2.41)	0.040	0.13 (0.26)	1.14 (0.68, 1.89)	0.617	0.67 (0.35)	1.95 (0.98, 3.88)	0.055
Hemoglobin	-0.68 (0.05)	0.51 (0.46, 0.56)	<0.001	-0.44 (0.22)	0.64 (0.42, 0.99)	0.045	-0.06 (0.07)	0.94 (0.82, 1.09)	0.454
BUN	0.03 (0.01)	1.03 (1.01, 1.05)	0.003	0.01 (0.01)	1.01 (0.99, 1.03)	0.317	-0.01 (0.01)	0.99 (0.97, 1.01)	0.381
-2(loglikelihood)	5061.89								
Association structure: both true slope and true value of longitudinal biomarkers ($\omega_{hk,ig}^3(t, \nu_{ig})$)									
Variables	Transition 1 to 2			Transition 2 to 3			Transition 1 to 3		
	Coef (SE)	HR (95% CI)	<i>p</i> -value	Coef (SE)	HR (95% CI)	<i>p</i> -value	Coef (SE)	HR (95% CI)	<i>p</i> -value
DRG	-0.01 (0.11)	0.99 (0.80, 1.23)	0.927	-0.03 (0.12)	0.97 (0.77, 1.23)	0.807	-0.13 (0.37)	0.88 (0.42, 1.82)	0.728
Recipient age	-0.015 (0.01)	0.98 (0.96, 1.00)	0.134	0.01 (0.005)	1.01 (1.00, 1.02)	0.024	0.01 (0.01)	1.01 (0.98, 1.04)	0.374
Creatinine	η_{11} 1.31 (0.38)	3.71 (1.76, 7.83)	<0.001	0.08 (0.32)	1.08 (0.59, 2.03)	0.802	1.44 (0.62)	4.22 (1.25, 14.29)	0.020
	η_{21} -1.85 (0.41)	0.16 (0.07, 0.35)	<0.001	0.54 (0.36)	1.72 (0.85, 3.47)	0.134	1.13 (0.59)	3.09 (0.97, 9.83)	0.055
Hemoglobin	η_{12} -0.83 (0.06)	0.44 (0.39, 0.49)	<0.001	-0.47 (0.07)	0.62 (0.54, 0.72)	<0.001	0.06 (0.09)	1.06 (0.89, 1.27)	0.484
	η_{22} -0.11 (0.07)	0.89 (0.78, 1.01)	0.083	-0.05 (0.08)	0.95 (0.81, 1.12)	0.571	0.43 (0.21)	1.54 (1.03, 2.31)	0.035
BUN	η_{13} 0.01 (0.01)	1.01 (0.99, 1.03)	0.074	0.01 (0.01)	1.01 (0.99, 1.02)	0.480	-0.01 (0.01)	0.98 (0.97, 1.00)	0.120
	η_{23} 0.05 (0.01)	1.05 (1.04, 1.06)	<0.001	0.03 (0.02)	1.03 (0.99, 1.07)	0.134	0.05 (0.03)	1.05 (0.99, 1.11)	0.095
-2(loglikelihood)	4981.07								
Association structure: the area under the biomarkers' trajectories ($\omega_{hk,ig}^4(t, \nu_{ig})$)									
Variables	Transition 1 to 2			Transition 2 to 3			Transition 1 to 3		
	Coef (SE)	HR (95% CI)	<i>p</i> -value	Coef (SE)	HR (95% CI)	<i>p</i> -value	Coef (SE)	HR (95% CI)	<i>p</i> -value
DRG	0.04 (0.11)	0.96 (0.78, 1.19)	0.742	0.26 (0.12)	1.30 (1.03, 1.64)	0.200	-0.47 (0.37)	0.62 (0.30, 1.28)	0.203
Recipient age	-0.05 (0.02)	0.95 (0.91, 0.99)	0.013	0.13 (0.05)	1.14 (1.03, 1.26)	0.009	0.01 (0.01)	1.01 (0.98, 1.04)	0.425
Creatinine	0.09 (0.04)	1.10 (1.01, 1.18)	0.024	0.07 (0.04)	1.07 (0.99, 1.16)	0.080	0.15 (0.08)	1.16 (0.99, 1.36)	0.061
Hemoglobin	-0.10 (0.04)	0.90 (0.84, 0.98)	0.012	-0.09 (0.04)	0.91 (0.84, 0.99)	0.023	-0.02 (0.01)	0.98 (0.97, 1.00)	0.134
BUN	0.12 (0.05)	1.13 (1.02, 1.24)	0.016	0.02 (0.01)	1.02 (1.00, 1.04)	0.069	-0.001 (0.001)	1.00 (0.99, 1.01)	0.133
-2(loglikelihood)	4013.425								

intensities between different states were the same in terms of the association direction and significance. On the other hand, the effect of incorporating the correlation between longitudinal biomarkers in the model was examined by comparing the use of univariate and multivariate prediction of biomarkers in the multistate model (Appendix A (Table A.2)). The results showed that the estimation of associations and their significant levels differ in two types of models. So, we can conclude that incorporating the correlation between longitudinal outcomes in the model is of particular importance. In fact, our proposed model is an extension of the two-stage model-based approach introduced by Guler et al. [8] for multivariate longitudinal and a single event of interest. Besides, we used multivariate linear mixed effects model to model the trajectories of longitudinal biomarkers, while they have used a pairwise approach as an alternative. Other features of our proposed model are similar to the model introduced by Guler et al., discussed in detail in [8].

In our case study, as it was expected, the level of creatinine and BUN can be used to monitoring post-transplant renal graft function. In fact, the levels of these biomarkers will increase in patients with a dysfunctional graft, and this increase is translated to a higher risk of rejection or mortality [20–23]. On the other hand, less hemoglobin level in the post-transplant period has been known to be associated with an increased risk of graft loss and mortality [24, 25] that was consistent with our findings. Moreover, some transplant medications are known to be associated with low levels of hemoglobin (anemia) [24], and the increased risk may be caused by these treatments. We showed that the prediction of measurements of investigated biomarkers at a fixed time point could be useful for monitoring the outcomes of transplantation. However, the ability of these measurements to predict the outcomes of interest is limited. Instead, here the cumulative effects of these biomarkers on the transplantations' outcomes are demonstrated. So, it is advised that the measurements of these biomarkers should be recorded in such a way that the calculation of cumulative effects is possible.

Data Availability

The datasets analyzed during the current study are available from the corresponding author on reasonable request.

Disclosure

This study has been adapted from the PhD thesis at the Hamadan University of Medical Sciences.

Conflicts of Interest

The authors declare that there are no conflicts of Interest.

Acknowledgments

The authors would like to thank the Vice-Chancellor of Research and Technology, Hamadan University of Medical Sciences, for the approval and support of the study. The study was funded by Vice-Chancellor for Research and Technology, Hamadan University of Medical Sciences (Grant no. 9812209742).

Supplementary Materials

Table A.1: results of comparison of joint models and two-stage based models for each longitudinal biomarker separately. Table A.2: results of comparison of two multistate models which included the estimation of the longitudinal biomarkers based on separated linear mixed effects models (Model 1) and a multivariate linear mixed effects model (Model 2) for longitudinal biomarkers. (*Supplementary Materials*)

References

- [1] D. Rizopoulos, *Joint Models for Longitudinal and Time-To-Event Data: With Applications in R*, Chapman and Hall/CRC, Boca Raton, FL, USA, 2012.
- [2] M. S. Wulfsohn and A. A. Tsiatis, "A joint model for survival and longitudinal data measured with error," *Biometrics*, vol. 53, pp. 330–339, 1997.
- [3] D. Zhang, M.-H. Chen, J. G. Ibrahim, M. E. Boye, and W. Shen, "Bayesian model assessment in joint modeling of longitudinal and survival data with applications to cancer clinical trials," *Journal of Computational and Graphical Statistics*, vol. 26, no. 1, pp. 121–133, 2017.
- [4] L. Ferrer, V. Rondeau, J. Dignam, T. Pickles, H. Jacqmin-Gadda, and C. Proust-Lima, "Joint modelling of longitudinal and multistate processes: application to clinical progressions in prostate cancer," *Statistics in Medicine*, vol. 35, no. 22, pp. 3933–3948, 2016.
- [5] Y.-Y. Chi and J. G. Ibrahim, "Joint models for multivariate longitudinal and multivariate survival data," *Biometrics*, vol. 62, no. 2, pp. 432–445, 2006.
- [6] E.-R. Andrinopoulou, D. Rizopoulos, J. J. M. Takkenberg, and E. Lesaffre, "Joint modeling of two longitudinal outcomes and competing risk data," *Statistics in Medicine*, vol. 33, no. 18, pp. 3167–3178, 2014.
- [7] E. Dantan, P. Joly, J.-F. Dartigues, and H. Jacqmin-Gadda, "Joint model with latent state for longitudinal and multistate data," *Biostatistics*, vol. 12, no. 4, pp. 723–736, 2011.
- [8] I. Guler, C. Faes, C. Cadarso-Suárez, L. Teixeira, A. Rodrigues, and D. Mendonça, "Two-stage model for multivariate longitudinal and survival data with application to nephrology research," *Biometrical Journal*, vol. 59, no. 6, pp. 1204–1220, 2017.
- [9] P. T. T. Huong, D. Nur, H. Pham, and A. Branford, "A modified two-stage approach for joint modelling of longitudinal and time-to-event data," *Journal of Statistical Computation and Simulation*, vol. 88, no. 17, pp. 3379–3398, 2018.
- [10] D. Hedeker and R. D. Gibbons, *Longitudinal Data Analysis*, Vol. 451, John Wiley & Sons, New York, NY, USA, 2006.

- [11] D. R. Cox, "Regression models and life-tables," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 34, no. 2, pp. 187–202, 1972.
- [12] W. Ye, X. Lin, and J. M. G. Taylor, "Semiparametric modeling of longitudinal measurements and time-to-event data—a two-stage regression calibration approach," *Biometrics*, vol. 64, no. 4, pp. 1238–1246, 2008.
- [13] Y. Pawitan and S. Self, "Modeling disease marker processes in AIDS," *Journal of the American Statistical Association*, vol. 88, no. 423, pp. 719–726, 1993.
- [14] A. A. Tsiatis, V. Degruttola, and M. S. Wulfsohn, "Modeling the relationship of survival to longitudinal data measured with error. Applications to survival and CD4 counts in patients with AIDS," *Journal of the American Statistical Association*, vol. 90, no. 429, pp. 27–37, 1995.
- [15] I. Guler, "Comparing the predictive performance of different regression models for longitudinal and time-to-event data," in *Workshop on Statistical Modelling*, F. S. T. Kneib and J. Fahrholz, Eds., pp. 111–116, Göttingen, Germany, 2014.
- [16] L. Wu, W. Liu, and X. J. Hu, "Joint inference on HIV viral dynamics and immune suppression in presence of measurement errors," *Biometrics*, vol. 66, no. 2, pp. 327–335, 2010.
- [17] R. M. Yücel, "R mlmm package: fitting multivariate linear mixed effects models with missing values," *Turkiye Klinikleri Journal of Biostatistics*, vol. 7, no. 1, 2015.
- [18] S. W. Lagakos, C. J. Sommer, and M. Zelen, "Semi-Markov models for partially censored data," *Biometrika*, vol. 65, no. 2, pp. 311–317, 1978.
- [19] S. Datta and G. A. Satten, "Validity of the Aalen-Johansen estimators of stage occupation probabilities and Nelson-Aalen estimators of integrated transition hazards for non-Markov models," *Statistics & Probability Letters*, vol. 55, no. 4, pp. 403–411, 2001.
- [20] I. D. Ha, T. Park, and Y. Lee, "Joint modelling of repeated measures and survival time data," *Biometrical Journal*, vol. 45, no. 6, pp. 647–658, 2003.
- [21] B. L. Kasiske, "Creatinine excretion after renal transplantation," *Transplantation*, vol. 48, no. 3, pp. 424–427, 1989.
- [22] K. D. Liu, J. Himmelfarb, E. Paganini et al., "Timing of initiation of dialysis in critically ill patients with acute kidney injury," *Clinical Journal of the American Society of Nephrology*, vol. 1, no. 5, pp. 915–919, 2006.
- [23] J. L. Lyman, "Blood urea nitrogen and creatinine," *Emergency Medicine Clinics of North America*, vol. 4, no. 2, pp. 223–233, 1986.
- [24] V. Ramanathan and S. Mandayam, "Post-transplant anemia," in *Management of Anemia*, pp. 185–198, Springer, New York, NY, USA, 2018.
- [25] A. Schechter, "Post renal transplant anemia: severity, causes and their association with graft and patient survival," *BMC Nephrology*, vol. 20, no. 1, p. 51, 2019.