The International Journal of Biostatistics

Volume 4, *Issue* 1 2008 *Article* 21

Joint Analysis of Current Status and Marker Data: An Extension of a Bivariate Threshold Model

Xingwei Tong, *Beijing Normal University*
Xin He, *The Ohio State University*
Jianguo Sun, *University of Missouri*
Mei-Ling T. Lee, *University of Maryland*

Recommended Citation:
Tong, Xingwei; He, Xin; Sun, Jianguo; and Lee, Mei-Ling T. (2008) "Joint Analysis of Current Status and Marker Data: An Extension of a Bivariate Threshold Model," *The International Journal of Biostatistics*: Vol. 4: Iss. 1, Article 21.
DOI: 10.2202/1557-4679.1122

Joint Analysis of Current Status and Marker Data: An Extension of a Bivariate Threshold Model

Xingwei Tong, Xin He, Jianguo Sun, and Mei-Ling T. Lee

Abstract

This paper considers joint analysis of current status and marker data using a threshold model based on first hitting times. A failure time is defined as the time at which a subject's latent health status process first decreases to zero. We extend the bivariate Wiener process model in Whitmore et al. (1998) to the case when only current status data are available. We develop maximum likelihood estimation procedures and provide simulation studies. We apply our methods to a motivating example involving liver tumors in mice.

KEYWORDS: bivariate Wiener process, current status data, first-hitting-time model, health status process, joint analysis, marker process

Author Notes: We wish to thank the editor and two referees for their thoughtful and constructive comments that greatly improved the article. We are also grateful to Dr. G. A. Whitmore for proofreading the final manuscript. This project was supported in part by NSFC 10671106 (Tong), NIH grants OH008649 (Lee, He) and EY12269 (Lee).

1 Introduction

Threshold regression, as reviewed in Lee and Whitmore (2006) and Aalen et al. (2008), is a relatively new methodology for analyzing time-to-event data and is based on a first-hitting-time model. In an engineering context, a mechanical failure can be defined as the time at which a degradation process first reaches a fixed threshold. Similarly, without assuming proportional hazards, a health failure occurs when a latent health status process first crosses a threshold or reaches a certain level. In both situations, it is often the case that there exists a time-varying longitudinal marker process that is related to and could be used to predict the latent process or failure time of interest. Whitmore et al. (1998) proposed a bivariate Wiener process as a joint model for both the latent process and the marker process. Lee et al. (2000) extended this bivariate model to include covariates. Interval-censored data or current status data, however, were not considered for these earlier threshold models in the literature and are the focus of this article.

The problem of joint modeling often occurs in medical studies in which there exist some markers or surrogates for a clinical disease end point or failure time of interest (Prentice, 1989; Jewell and Nielsen, 1993; Jewell and Kalbfleisch, 1996). Another common situation is the presence of time-dependent or longitudinal covariates, and the goal includes estimation of the effects of these covariates on the failure time of interest. One major difference between these conventional survival methods and threshold regression models based on first hitting times is that the markers or time-dependent covariates are usually observed completely or at different time points in the former while the marker process is often observed only at the first hitting or failure time in the latter. For the former situation, a number of models and inference procedures have been proposed in the literature when one observes right-censored or interval-censored failure time data (Kalbfleisch and Prentice, 2002; Sun, 2006). One of the most commonly used models is the proportional hazards model, which assumes that the markers or covariates affect the failure time in a proportional way.

With interval-censored failure time data, the failure time of interest is observed only to belong to some interval instead of being known exactly or to be greater than some censoring time as in right-censored failure time data (Sun, 2006). This is common in, for example, follow-up survival studies where each study subject is observed only periodically. Thus, the occurrence status of a failure is only known at these observation time points, resulting in the occurrence time lying in an interval, defined by the two observation time points immediately before and after the occurrence. A very important special case of interval-censored data is when each subject is observed only once and in this case, the failure time is only known to be smaller or greater than the observation time. Such data are commonly referred to as current status data and usually occur in, for example, demographical studies, tumorgenicity experiments, or general cross-sectional studies (Diamond et al., 1986; Jewell and Shiboski, 1990; Sun and Kalbfleisch, 1993). Van der Laan and Jewell (2003) investigated nonparametric estimation with current status and right-censored data on a counting process when a marker variable is observed at the censoring time.

An example of current status data, which motivated this study, is given in Section 5. The data were obtained from a 2-year tumorgenicity study conducted by the National Toxicology Program on the onset of various types of tumors in animals. Some animals died naturally during the study and those who survived at the end of study were sacrificed for examination of tumors. At death or sacrifice, animals were examined for the presence of various tumors in different organs. In this application, we only consider irreversible tumors and we wish to assess the predictive value of body weight of the animal as a marker process for the latent health status process that characterizes tumor onset. More details are given in Section 5.

The remainder of the paper is organized as follows. In Section 2, we begin by reviewing the basics of the bivariate Wiener model. We then generalize the model to the case of current status data and derive the corresponding likelihood function. For inference about unknown parameters, we develop the maximum likelihood approach in Section 3. The proposed approach is flexible in that it allows subjects to have different initial status and also it allows estimation of covariate effects. In Section 4 we present numerical simulation results conducted for the assessment of inference procedures that we developed. In Section 5, we apply the proposed methodology to the motivating tumorgenicity study discussed above. Section 6 contains some concluding remarks and discussion. Using a stochastic process to represent the latent health status, we will focus on current status data from medical studies in this article. The methodology developed in this article, however, is not limited to modeling health status and can be applied to first-hitting-time problems arising from other applications.

2 Bivariate Wiener Model and Likelihood Function for Current Status Data

Consider a medical study that involves two processes, one being an unobservable process $X(t)$ representing the latent health status of study subjects and the other a related marker process $Y(t)$ that is observable. Assume that $\{X(t), Y(t)\}$ follows a bivariate Wiener diffusion process with initial values $X(0)$ and $Y(0)$. That is, at any time t, the vector $\{X(t), Y(t)\}$ follows a bivariate normal distribution with mean $\{X(0), Y(0)\} + t\,\mu$ and covariance matrix $t\,\Sigma$, where $\mu = (\mu_x, \mu_y)$ and

$$\Sigma = \begin{pmatrix} \sigma_{xx} & \sigma_{xy} \\ \sigma_{yx} & \sigma_{yy} \end{pmatrix} .$$

The bivariate Wiener diffusion process has been studied by many authors, including Whitmore et al. (1998) and Lee et al. (2000), where the former considered the situation with $X(0) = 0$ fixed and the latter treated $X(0)$ as a positive unknown parameter. The unobservable latent health status process $X(t)$ can be modeled by borrowing strength from the observable marker process $Y(t)$. Let S denote the first time that $X(t)$ decreases to the failure threshold. That is, health degrades to the threshold at failure time S. Whitmore et al. (1998) and Lee et al. (2000) investigated situations where right-censored data are observed. They discussed cases where S is either observed exactly or known to be greater than a censoring time and for $Y(t)$ only its value at S or the censoring time is known.

Assume that a subject's initial health status is some unknown positive number denoted by $\delta = X(0)$. A failure is observed when the process $X(t)$ first hits 0, i.e., at the first-hitting-time S. Therefore, the random variable S, representing survival time, is defined as the time t at which $X(t) = 0$ with $X(s) > 0$ for all $0 < s < t$. Thus, failure occurs when the subject's health status decreases to the origin for the first time. For the marker process $Y(t)$, we assume that $Y(0) = 0$. Process standard deviations are denoted by $\sigma_x = \sigma_{xx}^{1/2}$ and $\sigma_y = \sigma_{yy}^{1/2}$. Furthermore, ρ is defined as the correlation parameter between $X(t)$ and $Y(t)$, that is,

$$\rho = \frac{\sigma_{xy}}{\sigma_x\,\sigma_y} .$$

Parameter ρ measures how closely the marker process $Y(t)$ can track or predict the latent health status process $X(t)$. In addition to process correlation coefficient ρ, the parameters of interest also include μ_x, the average change per unit time in health status. A negative value for μ_x would suggest that the subject's health status is decreasing and tending towards failure.

We now extend the model to include situations where only current status data are available from cross-sectional studies. Specifically, suppose that each subject is observed only once at time t_i and the observed data are

$$\{\,(t_i, \Delta_i \,=\, I(S_i \leq t_i), y_i \,=\, Y_i(t_i))\,;\, i \,=\, 1, ..., n\,\}\,,$$

where S_i and $Y_i(t)$ denote the failure time and marker process for subject i, respectively. To construct the likelihood function, let ϕ and Φ denote the density and distribution functions of the standard normal variable. For a subject with marker observation $Y(t) = y$, let $p_s(y)$ and $p_f(y)$ denote his/her probability density functions for surviving beyond time t or failing before time t, respectively. Following Lee et al. (2000), we can derive the function

$$\begin{aligned} p_s(y) \;&=\; \lim_{h\to 0} \frac{1}{h} Pr\{Y(t) \in (y, y+h), S > t\} \\ &=\; \Phi(c_1) q_1(y) \,-\, \exp\left(-\tfrac{2\delta\mu_x}{\sigma_x^2}\right) \Phi(c_2) q_2(y) \end{aligned}$$

and show that

$$\begin{aligned} p_f(y) \;&=\; \lim_{h\to 0} \frac{1}{h} Pr\{Y(t) \in (y, y+h), S \leq t\} \\ &=\; \lim_{h\to 0} \frac{1}{h} Pr\{Y(t) \in (y, y+h)\} - \lim_{h\to 0} \frac{1}{h} Pr\{Y(t) \in (y, y+h), S > t\} \\ &=\; q_1(y) \,-\, p_s(y) \\ &=\; q_1(y)\{1 - p_s(y) q_1^{-1}(y)\} \end{aligned}$$

where $q_1(y)$ is the marginal density function of $Y(t) = y$. Here $q_1(y)$ and $q_2(y)$ are density functions of normal distributions with means $\mu_y t$ and $\mu_y t - 2\delta\rho\sigma_y/\sigma_x$, respectively, with the same variance $\sigma_y^2 t$, and

$$c_1 \,=\, c_1(t) \,=\, \frac{\delta + \mu_{x,y}(t)}{\sigma_x \sqrt{(1-\rho^2)t}}\;,\; c_2 \,=\, c_2(t) \,=\, \frac{\delta + \mu_{x,*}(t)}{\sigma_x \sqrt{(1-\rho^2)t}}\,,$$

with

$$\mu_{x,y}(t) \,=\, \mu_x t + \rho\sigma_x\sigma_y^{-1}(y - \mu_y t)\,,\; \mu_{x,*}(t) \,=\, \mu_{x,y}(t) - 2\delta(1-\rho^2)\,.$$

Using the notation defined above, one can derive the likelihood function of the observed data $(t_i, \Delta_i, y_i), i = 1, \cdots, n$, as

$$\begin{aligned} L(\delta, \mu_x, \rho, \mu_y, \sigma_y, \sigma_x) \;&=\; \textstyle\prod_{i=1}^n [p_s(y_i)]^{1-\Delta_i} [p_f(y_i)]^{\Delta_i} \\ &=\; \textstyle\prod_{i=1}^n q_1(y_i) [p_s(y_i) q_1^{-1}(y_i)]^{1-\Delta_i} [1 - p_s(y_i) q_1^{-1}(y_i)]^{\Delta_i}\;. \end{aligned}$$

Thus the log-likelihood function of the observed data can be written in the form

$$l(\delta, \mu_x, \rho, \mu_y, \sigma_y, \sigma_x) = \sum_{i=1}^{n} \{\log[q_1(y_i)] + (1-\Delta_i)\log[p(y_i)] + \Delta_i \log[1-p(y_i)]\}, \tag{1}$$

where $p(y) = p_s(y)q_1^{-1}(y) = \Phi(c_1) - \Phi(c_2)\exp\{-(c_1^2 - c_2^2)/2\}$. In the next section, we will discuss estimation of unknown parameters and their inference.

3 MLE for Joint Modeling of Current Status and Marker Data

To estimate unknown parameters, it is natural to maximize the log-likelihood function (1). Similar to the right-censored data case considered in Lee et al. (2000), we found that, among all six parameters $(\delta, \mu_x, \rho, \mu_y, \sigma_y, \sigma_x)$, only five of them can be estimated. Because the process $X(t)$ is latent and therefore has an undefined scale, we have chosen to set $\sigma_x = 1$ and focus on estimation of $\theta = (\delta, \mu_x, \rho, \mu_y, \sigma_y)'$.

The maximum likelihood estimation of θ can be obtained by either solving the score function $U(\theta) = \partial l(\theta)/\partial\theta = 0$ or by maximizing the log-likelihood function directly with some constraints on the parameters. In this situation, the algorithm procedure can be simplified. Since the marginal density function of $Y(t) = y$ is the normal density function with mean $\mu_y t$ and variance $\sigma_y^2 t$, it can be easily shown that $\hat{\mu}_y = \sum_{i=1}^n y_i / \sum_{i=1}^n t_i$ and $\hat{\sigma}_y^2 = n^{-1}\sum_{i=1}^n \{(y_i - \hat{\mu}_y t_i)^2/t_i\}$. The other three estimators can be obtained by maximizing $l(\delta, \mu_x, \rho, \hat{\mu}_y, \hat{\sigma}_y)$ with the constraints $\delta > 0, \mu_x < 0$, and $|\rho| < 1$.

Let $\hat{\theta}$ denote the maximum likelihood estimate of θ defined above. It can be easily shown that $\sqrt{n}(\hat{\theta} - \theta_0)$ converges weakly to a normal random variable with mean zero and variance $I^{-1}(\theta_0)$, where θ_0 is the true value. The Fisher information $I(\theta_0)$ can be consistently estimated by $n^{-1}\sum_{i=1}^n U_i(\hat{\theta})U_i'(\hat{\theta})$ with $U_i(\theta) = \partial l_i(\theta)/\partial\theta$ and $l_i(\theta) = \log[q_1(y_i)] + (1-\Delta_i)\log[p(y_i)] + \Delta_i \log[1-p(y_i)]$. Specifically,

$$U_i(\theta) = \frac{\partial \log q_1(y_i)}{\partial\theta} + \left[\frac{\Delta_i}{p(y_i)} - \frac{1-\Delta_i}{1-p(y_i)}\right]\frac{\partial p(y_i)}{\partial\theta},$$

where

$$\frac{\partial \log q_1(y)}{\partial\theta} = -\left(0, 0, 0, 0, \frac{1}{\sigma_y}\right)' - \frac{y - \mu_y t}{\sigma_y^2 t}\left(0, 0, 0, -t, -\frac{y - \mu_y t}{\sigma_y}\right)'$$

and

$$\frac{\partial p(y)}{\partial \theta} = \phi(c_1)\left(\frac{\partial c_1}{\partial \theta} - \frac{\partial c_2}{\partial \theta}\right) + \exp\left(-\frac{c_1^2 - c_2^2}{2}\right)\Phi(c_2)\left(c_1\frac{\partial c_1}{\partial \theta} - c_2\frac{\partial c_2}{\partial \theta}\right),$$

with

$$\begin{aligned}\frac{\partial c_1}{\partial \theta} &= \frac{1}{\sqrt{(1-\rho^2)t}}\left(1, t, \frac{y-\mu_y t}{\sigma_y}, -\frac{\rho t}{\sigma_y}, -\frac{\rho(y-\mu_y t)}{\sigma_y^2}\right)' + \frac{\rho c_1}{\sqrt{1-\rho^2}}\,(0,0,1,0,0)', \\ \frac{\partial c_2}{\partial \theta} &= \frac{\partial c_1}{\partial \theta} - \frac{2}{\sqrt{(1-\rho^2)t}}\left(1-\rho^2, 0, -\delta\rho, 0, 0\right)'.\end{aligned}$$

For simplicity, we did not include covariates in the preceding developments. For medical applications, covariates are very important. Hence, we now consider cases where there is a vector of covariates Z that could affect the latent and marker processes through their effects on μ_x and μ_y. Assume that covariate effects can be described by

$$\mu_x = h_x(Z'\beta_x)\,,\ \mu_y = h_y(Z'\beta_y)\,, \tag{2}$$

where h_x and h_y are known link functions and β_x and β_y are unknown parameters representing the effects of the covariates on the latent and marker processes, respectively. For inference about β_x and β_y along with other parameters, one can develop similar maximum likelihood estimation procedures.

In time-to-event analysis, one would often treat markers as covariates and estimate the effects of all covariates on the disease or health status, or consider the marker as a surrogate for the health status and estimate the effects of covariates on the marker. Here we estimate the effects of covariates on the health status and marker separately. In model (2), β_x represents the direct effects of covariates on the health status, and β_y represents the indirect effects of covariates on the health status through the marker process.

Model (2) assumes that covariates Z have effects on μ_x and μ_y only. In practice, they may affect other parameters too or the effects may have different forms rather than the one specified by model (2). For these situations, similar inference methods can be easily developed.

4 Simulation Studies

In this section, we report results obtained from simulation studies conducted for assessing the performance of the developed methodology. In the study, we fixed $(\delta, \mu_x, \mu_y, \sigma_y) = (2, -1, 1, 2)$ and set $\rho = 0.05$, 0.25, 0.5 and 0.75. This gave $\mathrm{E}(S) = \delta/|\mu_x| = 2$, the expected failure time. Note that, as mentioned before, μ_x represents the mean change per unit time in health status, μ_y and σ_y describe the mean change and standard deviation per unit time of the marker process, and ρ measures the degree of correlation between the two processes.

For simulated data, we first generated observation times t_i's from a normal distribution with mean 2 and variance 0.25. To generate the latent process $X(t)$ and the marker process $Y(t)$, we took the approach of accumulating correlated normally distributed increments (Δ_x, Δ_y) over time increments of length $r = 0.01$. Specifically, for subject i, we first generated a simple random sample of size R_i $\{\Delta_{xj}, \Delta_{yj}\,;\, j = 1, ..., R_i\}$ from a bivariate normal distribution with mean $(r\,\mu_x, r\,\mu_y)$ and covariance matrix

$$V = \begin{pmatrix} 1 & 2\,\rho \\ 2\,\rho & 4 \end{pmatrix},$$

where $R_i = [t_i/r]$. Define $x_{ij} = \delta + \sum_{k=1}^{j} \Delta_{xk}$ and $y_{ij} = \sum_{k=1}^{j} \Delta_{yk}$, $j = 1, ..., R_i$, $i = 1, ..., n$. If $x_{ij} > 0$ for all j, we took $\Delta_i = 0$ and $y_i = y_{iR_i}$ for the sample (t_i, Δ_i, y_i) from subject i. That is, the subject has not failed yet at time t_i. If there exists j such that $x_{il} > 0$ for $l = 1, ..., j-1$ and $x_{ij} \leq 0$, then we defined $\Delta_i = 1$ and let $y_i = y_{ij} + u_i$, where u_i was generated from the normal distribution with mean $(t_i - rj)\,\mu_y$ and variance $(t_i - rj)\,\sigma_y^2$. That is, the subject has already failed before time t_i.

Table 1 presents simulation results on estimation of δ, μ_x, ρ, μ_y and σ_y for different values of ρ and n. The results were based on 1000 replications. The table gives the average biases of estimated parameters based on simulated data, the sample standard deviations of the estimates (SSDs), and the averages of estimated standard errors (ESEs). These results indicate that the proposed estimates seem to perform well for most situations and, especially, the estimated standard error is close to the sample standard deviation. One exception is the estimation of δ, for which larger sample sizes seem to be needed to obtain precise estimates.

The simulation results given above are for the case of no covariates. We also considered situations where covariates may exist. In particular, we investigated the case where there exists a treatment indicator Z with $Z = 0$ or 1 for half of the study subjects. For this situation, we assumed that μ_x and μ_y have the forms

$$\mu_x = \beta_{x0} + \beta_{x1}\,Z\,,\ \mu_y = \beta_{y0} + \beta_{y1}\,Z\,,$$

where β_{x0} and β_{y0} denote the mean changes for the subjects in the control group and β_{x1} and β_{y1} represent the treatment effects on the latent and marker

Table 1: Simulation results for estimated parameters.

		Sample size n=100			Sample size n=200		
ρ	Parameters	Bias	SSD	ESE	Bias	SSD	ESE
0.05	δ	0.1865	0.7360	0.7235	0.1134	0.4996	0.5013
	μ_x	-0.0621	0.4665	0.4511	-0.0252	0.3149	0.3174
	ρ	0.0020	0.1424	0.1409	-0.0010	0.0986	0.0972
	μ_y	-0.0022	0.1451	0.1399	-0.0017	0.1016	0.1000
	σ_y	-0.0100	0.1504	0.1401	-0.0010	0.1039	0.1007
0.25	δ	0.1383	0.7314	0.7182	0.1070	0.4867	0.4934
	μ_x	-0.0279	0.4561	0.4621	-0.0183	0.3055	0.3104
	ρ	0.0035	0.1319	0.1371	0.0005	0.0962	0.0943
	μ_y	-0.0024	0.1427	0.1439	0.0053	0.1003	0.1011
	σ_y	-0.0247	0.1411	0.1493	-0.0065	0.1033	0.1035
0.5	δ	0.1339	0.6917	0.6772	0.0905	0.4732	0.4666
	μ_x	-0.0312	0.4416	0.4265	-0.0078	0.2974	0.2949
	ρ	0.0018	0.1197	0.1178	0.0064	0.0825	0.0815
	μ_y	0.0046	0.1450	0.1447	-0.0023	0.0999	0.1011
	σ_y	-0.0077	0.1497	0.1387	-0.0090	0.1031	0.1031
0.75	δ	0.1079	0.5880	0.6181	0.0971	0.4148	0.4192
	μ_x	-0.0143	0.3747	0.3946	-0.0205	0.2628	0.2656
	ρ	0.0130	0.0827	0.0851	0.0020	0.0587	0.0588
	μ_y	0.0043	0.1416	0.1431	-0.0050	0.0966	0.1010
	σ_y	-0.0156	0.1388	0.1489	-0.0082	0.1006	0.1025

NOTE: SSD is the sample standard deviation of the estimates and ESE is the average of the estimated standard errors.

Table 2: Simulation results for estimated parameters with a binary covariate and $\rho = 0.5$.

	Sample size n=100			Sample size n=200		
Parameters	Bias	SSD	ESE	Bias	SSD	ESE
δ	0.1571	0.8961	0.9378	0.1064	0.6483	0.6327
β_{x0}	-0.0086	0.6052	0.6173	-0.011	0.4276	0.4081
β_{x1}	0.0194	0.3009	0.3163	0.0113	0.2166	0.2100
ρ	0.0275	0.1434	0.1357	0.0131	0.0886	0.0874
β_{y0}	-0.0171	0.1949	0.2074	0.0062	0.1447	0.1443
β_{y1}	0.0168	0.2518	0.2547	-0.0107	0.1747	0.1767
σ_y	-0.0221	0.1345	0.1517	-0.0093	0.1006	0.1036

NOTE: SSD is the sample standard deviation of the estimates and ESE is the average of the estimated standard errors.

processes, respectively. Using the same set-up as for Table 1 and setting $(\delta, \beta_{x0}, \beta_{x1}, \beta_{y0}, \beta_{y1}, \sigma_y) = (2, -1, 0.5, 1, 0.5, 2)$ and $\rho = 0.5$, representing a moderate correlation between $X(t)$ and $Y(t)$, we obtained the results presented in Table 2 for these parameters. They give similar conclusions as those given by Table 1.

5 An Illustrative Example

In this section we demonstrate the proposed method using a set of current status data arising from a 2-year tumorgenicity study conducted by the National Toxicology Program. The original study has male and female F344/N rats and B6C3F$_1$ mice exposed to chloroprene at concentrations of 0, 12.8, 32, or 80 ppm by inhalation, for 6 hours per day, 5 days per week, for 2 years. Some animals died naturally during the study and those who survived at the end of study were sacrificed for examination. The animals were examined at death or sacrifice for the presence of various irreversible tumors in different organs. In the study, in addition to the time to death or sacrifice and the tumor status at the time, the body weights of each animal were recorded at birth and at the death or sacrifice time. Body weight often provides a good marker for the health status of an animal. The animals in this study were fed under the same regime in each treatment group and thus it seems reasonable to use body weight of the animals to predict the health status characterized by tumor onset. In the following, we focus on 100 male and female B6C3F$_1$

mice in the 80 ppm chloroprene exposure group with respect to the occurrence of liver tumors.

To assess the predictive effect of body weight on the health status of the mice in the study, let $X(t)$ denote the health status of a mouse and $Y(t)$ the body weight of the mouse at day t, measured from birth. When $X(t)$ is decreasing towards 0, the liver cells are tending to undergo a malignant transformation which, when it occurs, is the initiation of a tumor. For convenience, we will take $Y(t)$ to be the body weight at day t minus the birth body weight and so we have $Y(0) = 0$. Also we let S denote time to the occurrence of a liver tumor. Define $Z = 1$ if the mouse was male and 0 otherwise. We assume that $X(t)$ and $Y(t)$ can be described by the models discussed in the previous sections and that μ_x and μ_y have the following forms

$$\mu_x = \beta_{x0} + \beta_{x1} Z \, , \; \mu_y = \beta_{y0} + \beta_{y1} Z \, ,$$

where β_{x0} and β_{y0} denote the mean changes of health status and the weight of female mice and β_{x1} and β_{y1} represent the gender effects on the latent health status process and the weight process, respectively.

Applying the inference procedures proposed in the previous sections, we obtained $\hat{\delta} = 29.7227$, $\hat{\beta}_{x0} = -0.0239$, $\hat{\beta}_{x1} = -0.0229$, $\hat{\rho} = 0.3735$, $\hat{\beta}_{y0} = 0.0424$, $\hat{\beta}_{y1} = -0.0190$, and $\hat{\sigma}_y = 0.3029$, with corresponding estimated standard errors 21.1651, 0.0472, 0.0120, 0.1637, 0.0019, 0.0027, and 0.0181, respectively. Except for δ, β_{x0} and β_{x1}, all other estimated parameters are significantly different from 0 with p-values smaller than 0.05. These results indicate that the health status of the female and male mice in the study were slowly declining from the initial health status of 29.7227 at the rates of 0.0239 and 0.0468 per day, respectively. In particular, the health status of female mice seems to decline slower than that of male mice. Note that δ, β_{x0} and β_{x1} are mildly significant, and one should be careful since the samples have few events and therefore may not be very informative about the parameters of the latent health status process. In terms of the predictive effect of body weight on health status, the results suggest that health status and body weight were significantly and positively correlated. That is, heavier mice seem to have longer tumor-free lives. The results given above also indicate that female mice seem to gain significantly more weight than male mice and thus enjoy longer tumor-free lives as mentioned above.

6 Concluding Remarks and Discussion

In the preceding sections, we developed inference procedures for threshold regression models for current status data. As mentioned in Whitmore et al. (1998), one advantage of this model is that it does not require the proportional hazards model assumption, which is commonly used for failure time data. Another major difference between the usual analysis of current status failure time data and the problem considered here is that for the former, the history of covariates is usually assumed to be known although the failure time of interest is not. For the latter case, as discussed in this article, the covariates are time-dependent variables and are observed only at a single observation time. Of course, longitudinal readings on the marker process, if available, could be used for inferences on the parameters of the marker process.

One direction for future research is to generalize the proposed method to situations in which general interval-censored failure time data are available. In this case, as mentioned above, the time at which the health status reaches its threshold is observed only to belong to an interval and the only information available for the marker process $Y(t)$ are its values at the left and right end points of the intervals. Although this generalization of the idea used here is similar, derivations of the score and information matrix, as well as the computations, will be much more complicated.

Another direction for future research is to incorporate semiparametric covariate functions in the proposed model. In this article, we assumed a linear functional form for the covariates. However, in other applications, a nonlinear function may be more appropriate and it would be helpful to develop methods for nonparametric estimation of the covariate function, for example, by using kernel or spline methods.

References

Aalen, O. O., Borgan, Ø., and Gjessing, H. K. (2008). *Survival and Event History Analysis: A Process Point of View.* New York: Springer.

Diamond, I. D., McDonald, J. W., and Shah, I. H. (1986). Proportional hazards models for current status data: Application to the study of differentials in age at weaning in Pakistan. *Demography* **23**, 607-620.

Jewell, N. P. and Kalbfleisch, J. D. (1996). Marker processes in survival analysis. *Lifetime Data Analysis* **2**, 15-29.

Jewell, N. P. and Nielsen, J. P. (1993). A framework for consistent prediction rules based on markers. *Biometrika* **80**, 153-164.

Jewell, N. P. and Shiboski, S. C. (1990). Statistical analysis of HIV infectivity based on partner studies. *Biometrics* **46**, 1133-1150.

Kalbfleisch, J. D. and Prentice, R. L. (2002). *The Statistical Analysis of Failure Time Data*. Second edition, New York: Wiley.

Lee, M.-L. T., DeGruttola, V., and Schoenfeld, D. (2000). A model for markers and latent health status. *Journal of the Royal Statistical Society, Series B* **62**, 747-762.

Lee, M.-L. T. and Whitmore, G. A. (2006). Threshold regression for survival analysis: modeling event times by a stochastic process reaching a boundary. *Statistical Sciences* **21**, 501-513.

Prentice, R. L. (1989). Surrogate endpoints in clinical trials: definition and operational criteria. *Statistics in Medicine* **8**, 431-440.

Sun, J. (2006). *The Statistical Analysis of Interval-censored Failure Time Data*. New York: Springer.

Sun, J. and Kalbfleisch, J. D. (1993). The analysis of current status data on point processes. *Journal of the American Statistical Association* **88**, 1449-1454.

van der Laan, M. J. and Jewell, N. P. (2003). Current status and right-censored data structures when observing a marker at the censoring time. *The Annals of Statistics* **31**, 512-535.

Whitmore, G. A., Crowder, M. J., and Lawless J. F. (1998). Failure inference from a marker process based on a bivariate Wiener model. *Lifetime Data Analysis* **4**, 229-251.