Journal of Biopharmaceutical Statistics

ISSN: 1054-3406 (Print) 1520-5711 (Online) Journal homepage: http://www.tandfonline.com/loi/lbps20

# A Threshold Regression Mixture Model for Assessing Treatment Efficacy in a Multiple Myeloma Clinical Trial

Mei-Ling Ting Lee , Mark Chang & G. A. Whitmore

*Journal of Biopharmaceutical Statistics*, 18: 1136–1149, 2008

ISSN: 1054-3406 print/1520-5711 online
DOI: 10.1080/10543400802398524

# A THRESHOLD REGRESSION MIXTURE MODEL FOR ASSESSING TREATMENT EFFICACY IN A MULTIPLE MYELOMA CLINICAL TRIAL

**Mei-Ling Ting Lee[1], Mark Chang[2], and G. A. Whitmore[3]**
[1]*Department of Epidemiology and Biostatistics, School of Public Health, University of Maryland, College Park, Maryland, USA*
[2]*Millennium Pharmaceuticals, Inc., Cambridge, Massachusetts, USA*
[3]*McGill University, Montreal, Canada*

***A first-hitting-time (FHT) survival model postulates a health status process for a patient that gradually declines until the patient dies when the level first reaches a critical threshold. Threshold regression (TR) is a new regression methodology that incorporates the effects of covariates on the threshold and process parameters of this FHT model. In this study, we use TR to analyze data from a randomized clinical trial of treatment for multiple myeloma. The trial compares VELCADE and high-dose Dexamethasone, the former a new therapy and the latter an established therapy for this disease. Patients are switched between the two drugs based on patient response. The novel contribution of this work is the modeling of this clinical trial design using a mixture of TR models. Specifically, we propose a mixture FHT model to fit the survival distribution. The model includes a composite time scale that differentiates the rate of disease progression before and after switching. The analysis shows significant benefit from initial treatment by VELCADE. A comparison is made with a Cox proportional hazards regression analysis of the same data.***

***Key Words:*** Analysis time; Composite time; Cox regression; Data analysis; Disease progression; Duration; First hitting time; Lifetime; Maximum likelihood; Mixture model; Multiple myeloma; Proportional hazards; Randomized clinical trial; Stochastic process; Stopping time; Survival; Threshold regression; Time-to-event; Treatment switching; Wiener diffusion process.

## INTRODUCTION

Survival analysis in a clinical trial with treatment switching is a challenge. In randomized oncology trials, a patient's treatment may be switched in the middle of the study because a marker, such as disease progression (PD), indicates failure of the initial treatment regimen. In such cases, the total survival time from randomization is the sum of two event times, randomization-to-switch and switch-to-death. If the test drug is more effective than the control, then the majority of patients in the control group will switch to the test drug, and the survival difference between the two treatment groups will be significantly reduced in comparison to

Received June 26, 2007; Accepted February 4, 2008
Address correspondence to Mei-Ling Ting Lee, Department of Epidemiology and Biostatistics, School of Public Health, University of Maryland, College Park, MD 20742, USA; E-mail: MLTLEE@UMD.edu

the case without treatment switching. This marker-based treatment switch is not random switching but rather response-adaptive switching. As a result of treatment switching, the treatment effect can only be partially observed, and the effects of different treatments are difficult to distinguish. Branson and Whitehead (2002) used latent event times to model a patient's observed survival time after switching and the survival time that would have been observed if this patient had not switched treatment. Their model, however, does not take into account the fact that a treatment switch is often based on the observed effect of the current treatment. For example, the survival time of a patient who switches from the active control to the test treatment might be longer than his survival time if he had adhered to the original treatment. Therefore, the Branson and Whitehead model is one for random treatment switching with a constant latent hazard rate over time. In fact, even in the case of random switching, the hazard rate increases after the switch, and the later the switch occurs, the larger the hazard rate after the switch. Shao et al. (2005) proposed a generalized time-dependent Cox's proportional hazard model and provide maximum likelihood estimates (MLE) of the parameters. However, the method for hypothesis testing was not provided.

In this study, we use a flexible mixture of threshold regression models to approach this issue. Specifically, we employ a mixture of two first-hitting-time (FHT) distributions for an underlying Wiener process representing patient health status. This kind of mixture model has been used in other health studies (Balka, 2005; Whitmore and Su, 2007) but not to model data from a randomized clinical trial. The mixture model includes a composite time scale that differentiates between the rate of disease progression during the interval from randomization to switching and the interval from switching until death. The proposed model effectively and realistically captures the nonlinear complex nature of the survival process. The proposed model is applied to an oncology trial for patients with multiple myeloma. The results show that the model works well in dealing with treatment switching and gives new and significant insights into the disease process and therapeutic effects.

Cox proportional hazards regression is the conventional model of choice for studies of censored survival data with covariates. The assumption of proportional hazards, however, is not always appropriate. In addition, the estimated baseline hazard function in this model is a nonparametric byproduct that is rarely an object of study in its own right, in part because its shape offers investigators few scientific insights. Many investigators, such as Royston and Parmar (2002), have proposed alternatives to the Cox regression model that attempt to address these weaknesses. Threshold regression (TR) models are parametric models that do not generally possess the proportional hazards property, although variants exist that have this property if it is deemed appropriate. Threshold regression models always require an explicit formulation of the underlying health process and the triggers that define event times. These strengths, its flexibility, and its ease of estimation have led us to adopt TR for our study here.

## STUDY PROTOCOL AND DATA

VELCADE (bortezomib) for injection is a FDA-approved drug for treatment of patients with multiple myeloma who have received at least one prior therapy. The pivotal study was a multi-center, international, randomized, open-label study designed to determine whether treatment with VELCADE 1.3 mg/m2/dose prolongs time-to-progression (TTP) relative to treatment with high-dose Dexamethasone in patients with multiple myeloma who require second-, third-, or fourth-line treatment. Patients were assigned to receive VELCADE or high-dose Dexamethasone by random allocation at a 1:1 ratio. Randomization was stratified based on the number of treatment regimens the patient previously received (one previous treatment regimen versus more than one previous treatment regimen) and stem cell transplant history (history of transplant versus no history of transplant). Patients were evaluated at scheduled visits in four study periods, Pre-treatment, Treatment, Short-term Follow-up, and Long-term Follow-up. All patients attended an End of Treatment visit 30 days after the last study drug dose. (For patients assigned to Dexamethasone, the End of Treatment visit could coincide with the week-36 "on treatment" visit.) After the End of Treatment visit, patients who had not experienced confirmed PD attended Short-term Follow-up visits every 6 weeks until development of confirmed PD. After development of PD, patients were followed for survival during the Long-term Follow-up Period every 3 months via telephone or office visit. Patients who developed confirmed PD after receiving Dexamethasone in the current study were eligible to receive VELCADE in another MPI-sponsored clinical study. Patients who developed confirmed PD after receiving VELCADE could switch to other treatments suggested by their clinicians. In most cases the alternate treatment included Dexamethasone. In our analysis, in fact, we assume that Dexamethasone was chosen as the alternate to VELCADE.

The primary endpoint was TTP, but the secondary endpoints included survival, quality of life, complete response, and partial response. A single interim analysis was conducted for TTP, based on the method of O'Brien and Fleming (1979). This test was performed when about half of the anticipated number of events had occurred (a total of 196 patients having PD). Statistical significance was to be declared at the interim analysis if the log-rank $P$-value was smaller than 0.005, and, failing this, at the final analysis if the $P$-value was smaller than 0.048. It was calculated that the study design had 80% power to detect a hazard ratio of 1.775. This corresponds to a median TTP for VELCADE of 14.2 months relative to 8.0 months for high-dose Dexamethasone. The resulting sample size was 310 patients per group. The log-rank test is a standard test for time-to-event analyses for New Drug Applications (NDA) submissions to the FDA and was originally used for the analysis.

## THRESHOLD REGRESSION (TR)

### FHT Models

Many forms of time-to-event data may be interpreted as the elapsed time from initial observation until the sample path of a parent stochastic process $\{X(t), t \geq 0\}$ first encounters a boundary set $\mathscr{B}$. For example, a lethal pancreatic cancer progresses toward a critical state that proves fatal for the patient. The elapsed time from initial diagnosis to the critical state is an FHT. A patient with progressive kidney disease eventually experiences renal failure. The elapsed time from recruitment to renal failure is an FHT. Many different mathematical forms may be used for the parent stochastic process that describes the progress of a disease and for the boundary set that will trigger the hitting time. The flexibility in choosing the parent process and boundary set give FHT models great scope for describing real medical and health situations. They are also appealing because of their conceptual unity, realism, and ease of application. Consequently, FHT models find many potential applications in health and medicine, as well as other fields, such as engineering and the social sciences.

In many applications, including our application here, there is no opportunity to observe the underlying parent stochastic process or the boundary set directly. Yet time-to-event and covariate data do provide some insight into the nature of these latent or unobservable characteristics of an FHT statistical model, as our application will demonstrate. The latency of the postulated health status process may seem to introduce a vague element to the model, but it is important to keep in mind that it is well defined clinically. Clinicians dealing with multiple myeloma patients are readily able to assess the health status of their patients. Our latent health status process may be considered as a one-dimensional index of how the patient's status changes through time under treatment. As described in Lee and Whitmore (2006), linear combinations of observable covariate processes can emulate a latent parent health status process and thus allow it to be monitored. In this study, the beta-2 microglobulin level of a patient, which we use as a baseline covariate, is one such observable covariate process. We show later in our results section that it is highly correlated with the health status of a multiple myeloma patient. Threshold regression forces an explicit formulation of the latent process and its observable associated processes and thereby encourages a closer study of the phenomenon.

In this application, we use a Wiener diffusion process to describe the health status of a subject with multiple myeloma. The meandering sample path of a Wiener process seems suited to describing the random ups and downs of the health status of a multiple myeloma patient. The Wiener process has a mean parameter $\mu$ and a variance parameter $\sigma^2$. We fix the boundary at zero so death is triggered when the subject's health status first reaches the zero level. We take the initial level of the process at time $t = 0$ to be positive, i.e., $X(0) = x_0 > 0$. We denote the FHT by $S$ and observe that in this theoretical setting $S$ has an inverse Gaussian distribution (Chhikara and Folks, 1989; Cox and Miller, 1965).

There is no guarantee that a parent stochastic process will reach a boundary set, and this is a possibility that we will face in our multiple myeloma study. This possibility may be described as the *cure rate* of the FHT model and we denote it here by $P(\text{cure})$. In the case of a Wiener process, starting at $x_0 > 0$ and having a boundary at zero, there is a positive probability that the sample path will not reach the boundary if its sample paths tend to drift away from the zero level (i.e., if its mean parameter $\mu$ is positive). This probability is given by

$$P(\text{cure}) = 1 - \exp(-2x_0\mu/\sigma^2) \tag{1}$$

As our study involves a latent health status process, the health status scale has an arbitrary unit of measurement. We therefore have one redundant parameter. In this report, we choose to set the variance parameter of the Wiener diffusion process to unity, i.e., to set $\sigma^2 = 1$. The implication is that both the initial health level $x_0$ and the mean parameter $\mu$ of the health status process will be measured in units of the standard deviation of the process.

### Basics of Threshold Regression

First-hitting-time models must usually include regression structures in order to capture the reality of practical applications. Regression structures allow covariates to account for variability in the data and thereby sharpen statistical inferences. Regression structures also provide scientific insights into the roles of covariates in the causal structure of underlying processes, boundary sets, and time scales. An FHT model with a regression structure is referred to as a TR model. The word *threshold* refers to the fact that the first hitting time occurs when the underlying process reaches a threshold state within a boundary set.

In TR, parameters of underlying processes, boundary sets, and time scales are connected to linear combinations of covariates using suitable regression link functions, as illustrated next for an arbitrary parameter $\zeta$.

$$g_\zeta(\zeta_i) = \boldsymbol{z}_i\boldsymbol{\beta} \tag{2}$$

Here $g_\zeta$ is the link function, parameter $\zeta_i$ is the value of parameter $\zeta$ for individual $i$, $\boldsymbol{z}_i = (1, z_{i1}, \ldots, z_{ik})$ is the covariate vector of individual $i$ (with a leading unit to include an intercept term), and $\beta$ is the associated vector of regression coefficients. The mathematical form of the link function must be suited to the application. Generally, it will be chosen to map the parameter space into the real line. For example, in our application, the initial health status parameter $x_0$ is linked to a linear combination of covariates using the natural logarithmic link function $\ln(x_0)$. On the other hand, the mean parameter $\mu$ is simply given an identity link. As is the case in conventional regression analysis, TR requires judicious choices of the link function for each parameter, the list of covariates entering the regression function, and the exact mathematical forms of the covariates in the regression function.

For additional background on first hitting time models and TR, the reader is referred to the following readings. Lee et al. (2000) use a bivariate Wiener diffusion process as the basis of a regression model for the study of progression to death in AIDS, with CD4 cell count serving as a marker process. Aalen and Gjessing (2001, 2004) provide excellent discussions of the importance of analyzing survival data from the point of view of a stochastic process. Lawless (2003) gives a summary of theory, models, and methods on process-defined failure models (see Section 11.5, pp. 518–523). Horrocks and Thompson (2004) model event times with multiple outcomes using the Wiener Process with drift. Padgett and Tomlinson (2004, 2005) investigate inference from accelerated degradation and failure data based on Gaussian process models, geometric Brownian motion, and gamma processes. Lee and Whitmore (2006) provide an overview of FHT models for survival and time-to-event data and a survey of threshold regression models.

### Threshold Regression Setup for This Study

Figure 1 gives the basic setup of the FHT model in this study. The horizontal scale is calendar time. The vertical scale represents multiple myeloma health status

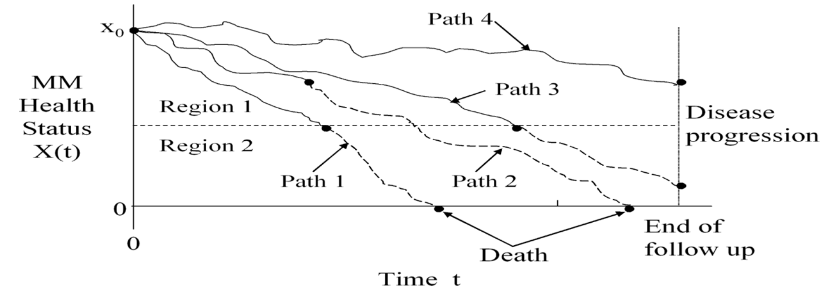


**Figure 1** The FHT model for multiple myeloma health status and time to PD and death. The figure shows four illustrative sample paths. Path 1 is a patient who experiences PD under the assigned primary therapy and then dies before the end of follow-up. The dashed sample path indicates assignment to the alternate therapy at PD. Path 2 is a patient who is switched to the alternate therapy prior to PD and happens to die before the end of follow-up. Path 3 is a patient who experiences PD on the primary therapy, is switched to the alternate therapy and then is living at the end of follow-up. Path 4 is a patient who does not experience PD under the primary therapy.

and the zero level on this scale represents death. The vertical scale is divided into two regions. The first region describes health status after randomization until the patient experiences PD. The second region describes health status between PD and death. The setup assumes that PD always precedes death. Patient records in this study are subject to censoring for both PD and death. The figure shows four illustrative sample paths for a multiple myeloma patient participating in the study. Path 1 illustrates a patient who experiences PD under the assigned primary therapy and then dies before the end of follow-up. The dashed sample path indicates that the patient is assigned to the alternate therapy at PD. Time of PD and time of death are both known in this case. Path 2 illustrates a patient who is switched to the alternate therapy prior to PD and happens to die before the end of follow-up. The time of PD is not observed in this case but the time of death is known. Path 3 illustrates a patient who experiences PD on the primary therapy, is switched to the alternate therapy, and then is living at the end of follow-up. The time of PD is known in this case, but survival time is censored. Path 4 is a patient who does not experience PD under the primary therapy. This patient is censored for both times of PD and death. In all cases, we note that the time on the primary therapy and the time on the alternate therapy are both measured, with the time measurement ending at death or end of follow up as the case may be. A feature not represented in Fig. 1 is a difference between the administrative censoring dates for PD and survival in this study—the administrative censoring date for survival follows one month after the administrative censoring date for PD.

### Regression Covariates and Treatment Regimens

The covariates used in the TR analysis are as follows.

1. *trt*—Indicator variable for the assigned primary treatment (−1, Dexamethasone; 1, VELCADE).
2. *pd*—Indicator variable for whether PD is observed under the primary therapy (−1, PD censored; 1, PD observed).

3. *trt_pd*—Interaction term representing a product of the indicator variables *trt* and *pd*.
4. *prev*—Indicator variable for number of previous treatments (0, one previous treatment; 1, more than one previous treatment).
5. *lgb2*—Baseline level of beta-2 microglobulin (the natural logarithm of the level, expressed in grams per liter).
6. *age*—Baseline age (in years).

The covariate *lgb2* is included because it is known to reflect the tumor burden of the patient. The indicator variables for *trt* and *pd* have been coded $-1$ and 1 to reduce multicollinearity with the interaction term *trt_pd*.

The four possible combinations of *trt* and *pd* define four treatment regimens for patients. A patient is assigned one of the primary therapies (Dexamethasone or VELCADE) and either experiences PD under the primary therapy or not, before proceeding to the alternate therapy. Thus, two time measurements are available for each patient, denoted by *dur_pd* and *post_pd*. These measurements represent time on the primary therapy and time on the alternate therapy, respectively. The latter could be zero (in the situation illustrated by path 4 in Fig. 1). The sum of these two measurements for each patient is the survival time or censored survival time, as the case may be. We denote this variable by *survdays*. We use the indicator variable *fail* to denote whether the survival time is observed or censored.

### Analysis Using a Composite Time Scale

Many authors have recognized that survival data may need to be modeled using time scales other than calendar time. See, for example, Oakes (1995), Duchesne and Lawless (2000), and Lawless (2003). Depending on the context, these alternative time scales have been given different names, such as analysis time, operational time, composite time, running time, and so on. The label *composite time* suits our application. We have already noted that the actual or censored survival time in our study is composed of two intervals, representing the time on the primary therapy (*dur_pd*) and the time on the alternative therapy (*post_pd*). In recognition of the fact that multiple myeloma may progress at different rates in these two intervals (irrespective of the treatment), we transform survival times from calendar time to a composite time scale, which has the following form.

$$r = \alpha t_1 + t_2 \tag{3}$$

Here $t_1$ and $t_2$ correspond to *dur_pd* and *post_pd*, respectively. Parameter $\alpha$ is the ratio of the rate of progression on the primary therapy relative to that on the alternate therapy. We use $r$ to denote the composite time scale.

## FIRST-HITTING-TIME MIXTURE MODEL

### Distribution Forms

As noted earlier, the FHT distribution for a boundary by a Wiener diffusion model, as illustrated in Fig. 1, follows an inverse Gaussian survival distribution.

This distribution depends on the initial health status level ($x_0$) and the mean and variance parameters ($\mu$ and $\sigma^2$) of the underlying Wiener process. We let $f(r \mid \mu, \sigma^2, x_0)$ and $F(r \mid \mu, \sigma^2, x_0)$ denote the probability density function (p.d.f.) and cumulative distribution function (c.d.f.) of the FHT distribution, both defined in terms of the composite time scale $r$. These functions have simple computational forms. The p.d.f. for the FHT is given by

$$f(r \mid \mu, \sigma^2, x_0) = \frac{x_0}{\sqrt{2\pi\sigma^2 r^3}} \exp\left[-\frac{(x_0 + \mu r)^2}{2\sigma^2 r}\right], \quad \text{for } -\infty < \mu < \infty, \ \sigma^2 > 0, \ x_0 > 0 \tag{4}$$

If $\mu > 0$ then the FHT is not certain to occur and the p.d.f. is improper. Specifically, the probability that the boundary will not be reached is given by the cure-rate expression in (1). The c.d.f. corresponding to (4) is

$$F(r \mid \mu, \sigma^2, x_0) = \Phi\left[-\frac{(\mu r + x_0)}{\sqrt{\sigma^2 r}}\right] + \exp(-2x_0\mu/\sigma^2)\Phi\left[\frac{\mu r - x_0}{\sqrt{\sigma^2 r}}\right] \tag{5}$$

where $\Phi(\cdot)$ is the c.d.f. of the standard normal distribution. The corresponding survival function is given by $\overline{F}(r \mid \mu, \sigma^2, x_0) = 1 - F(r \mid \mu, \sigma^2, x_0)$.

The forms of the p.d.f and c.d.f. in (4) and (5) show clearly that parameters $\mu$, $x_0$ and $\sigma^2$ are not mutually estimable from censored survival data. One of these parameters must be fixed. As noted earlier, we have arbitrarily set $\sigma^2$ equal to one.

### Mixture Model

We first summarize the overall survival results in this clinical trial setting. Figure 2 shows a comparison of Kaplan–Meier (KM) survival distributions for the two treatments, plotted against survival time measured in days from randomization.

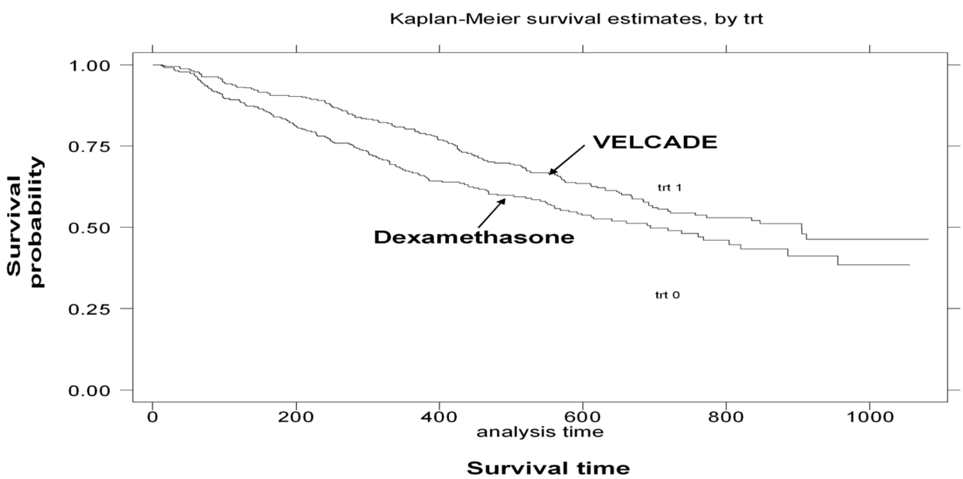


**Figure 2** Comparison of Kaplan–Meier survival distribution plots for the two treatments suggests VELCADE, as the primary therapy, offers better survival prospects than Dexamethasone.

The plot suggests that VELCADE increases survival time when it is the primary therapy. The plot by itself, however, is not based on a parametric model and does not show the dependence of survival prospects on other covariates. The plot simply suggests that VELCADE, as a primary treatment, offers a survival benefit but provides no insight into the source or nature of the benefit. As a preliminary quantification of the survival benefit of VELCADE, we report that a conventional Cox proportional hazards regression analysis with the treatment indicator variable *trt* as the only covariate gives a hazard ratio of .7259 ($P$-value $= 0.010$) for the two treatments. Recall that the *trt* variable is coded 1 and $-1$ rather than the conventional 1 and 0.

Although the inverse Gaussian FHT model is a plausible parametric model for the survival distribution of a single subject, the plotted survival curves suggest that the survival patterns across subjects in this study may be a mixture of two different inverse Gaussian distributions. The Kaplan–Meier survival plot for each primary treatment shows an initial plateau of low risk and then a secondary slackening of risk about 1 year out. Histograms of the survival times for each treatment also show diffused frequency patterns that are suggestive of overlapping modes. Our subsequent modeling also shows that, even after taking account of the covariates, there is significant statistical evidence of two components. The data are not adequate to tell if there are further minor mixture components. Denoting the survival function of the mixture model by $\overline{G}$, we have

$$\overline{G}(r) = p\overline{F}_1(r) + (1-p)\overline{F}_2(r) \tag{6}$$

The mixing parameter is proportion $p$. The $\overline{F}_j(r), j = 1, 2,$ are the respective component survival functions of the mixture. All the survival functions are expressed in terms of composite time as defined in Equation (3). We note that by defining the survival functions in terms of composite time, we assist in making the assumption of a fixed mean parameter for each component survival function more plausible.

### Statistical Inference

Each component survival function of the mixture model (6) has its own p.d.f. $f_j(r)$ and c.d.f. $F_j(r)$, as well as its own initial health status $x_{0j}$ and mean parameter $\mu_j$, for $j = 1, 2$. For compactness, we now denote the vector of parameters of our mixture model by $\boldsymbol{\theta}$. This vector includes all of the regression coefficients for parameters $p, \mu_1, x_{01}, \mu_2, x_{02}$, and $\alpha$, where the last one is the rate parameter in the composite time formula (3). We let $r_i$ denote the composite time for patient $i$. Time $r_i$ is the composite time at death for a dying patient and a right-censored composite time for death for a surviving patient. Hence, each dying patient $i$ contributes probability density $g(r_i \mid \boldsymbol{\theta})$ to the sample likelihood function, for $i = 1, \ldots, n_1$, where $g(r \mid \boldsymbol{\theta})$ is the mixture p.d.f. corresponding to model (6), and $n_1$ is the number of patients who die before the end of the study. In addition, each surviving patient $i$ contributes survival probability $\overline{G}(r_i \mid \boldsymbol{\theta})$ to the sample likelihood function, for $i = n_1 + 1, \ldots, n_1 + n_0$, where $\overline{G}(r \mid \boldsymbol{\theta})$ is the mixture survival function in (6), and $n_0$ is the number of patients who survive to the end of the study. We assume that this censoring is uninformative. The sum $n = n_1 + n_0$ is the total number of patients. The sample log-likelihood function to be maximized therefore has the form

$$\ln L(\boldsymbol{\theta}) = \sum_{i=1}^{n_1} \ln g(r_i \mid \boldsymbol{\theta}) + \sum_{i=n_1+1}^{n_1+n_0} \ln \overline{G}(r_i \mid \boldsymbol{\theta}) \tag{7}$$

We use the numerical gradient optimization routine *lf* in *Stata* to find the maximum likelihood estimate of the regression coefficient vector $\boldsymbol{\theta}$. This optimization routine requires only the sample log-likelihood function in (7). Derivatives of the likelihood function are not required.

## ANALYSIS AND RESULTS FOR THE MIXTURE MODEL

The data were prescreened to eliminate 3 subjects with zero survival times (*survdays* = 0), a further 50 subjects who were censored for PD at the outset of the study (*dur_pd* = 0), and a further 13 cases with missing beta-2 microglobulin readings. These eliminations left 603 subjects in the study.

The linear mixture of two inverse Gaussian first-hitting-time distributions in (6) has been fitted to the survival data. In the TR analysis, we choose the log-link function for initial health status parameters $x_{01}$ and $x_{02}$ and for the composite time parameter $\alpha$, an identity link function for the mean parameters $\mu_1$ and $\mu_2$, and a logit link function for the mixture probability parameter $p$.

Table 1 shows the regression output for this TR model. The parameter estimates are exact maximum likelihood estimates. The reported standard errors are asymptotic estimates derived numerically by the Stata routine from the negative inverse of the Hessian matrix for the sample likelihood function. The initial health status parameters have been made to depend on covariates *prev*, *lgb2*, and *age*, all of which are known at baseline. The mean parameters have been made to depend on covariates *trt*, *pd*, and their interaction term *trt_pd*. These covariates define the treatment regimen. The mixing parameter $p$ and composite time parameter $\alpha$ are not made to depend on covariates, although we revisit this point a little later. In interpreting the regression results, we use a $P$-value cutoff for each coefficient that is more stringent than the usual 0.05 value because we are estimating 18 regression coefficients. Specifically, we focus on effects that have $P$-values less than 0.003, where $0.05/18 = 0.003$. This adjustment represents a Bonferroni correction.

The first component of the survival mixture model (index $j = 1$) represents an estimated 14% of the survival probability because $\hat{p} = 0.14$. An estimated 95 percent confidence interval for $p$ is $0.09 \leq p \leq 0.21$. Thus, the two-component mixture model offers a significantly larger likelihood than an unmixed model, confirming that a mixture model is needed here. For the survival function of the first component, which is $F_1(r)$ in (6), the initial health status parameter $x_{01}$ depends significantly only on *age*, and even this significance is borderline at 0.003. The positive coefficient suggests that survival prospects from this component are a little more favorable for older patients, with the initial health status parameter $x_{01}$ rising by about 2% with each year of age. The mean parameter for the first survival component $\mu_1$ shows a strong treatment effect favoring VELCADE as the primary therapy (i.e., when *trt* equals 1). Taking account of the interaction effect, the relative

**Table 1** Threshold regression output for multiple myeloma survival time, based on the inverse Gaussian mixture model with composite time (6)

| Parameter | Variable | Estimate | Std. Error | $P$-value |
|---|---|---|---|---|
| $\ln(x_{01})$ | *prev* | −.20819 | .17817 | 0.243 |
| | *lgb2* | −.11418 | .12578 | 0.364 |
| | *age* | .02092 | .00695 | 0.003 |
| | constant | 1.11452 | .44204 | 0.012 |
| $\mu_1$ | *trt* | .12136 | .02659 | 0.000 |
| | *pd* | .07605 | .02722 | 0.005 |
| | *trt_pd* | −.04951 | .02655 | 0.062 |
| | constant | −.07338 | .03804 | 0.054 |
| $\ln(x_{02})$ | *prev* | −.34981 | .08117 | 0.000 |
| | *lgb2* | −.46291 | .06619 | 0.000 |
| | *age* | −.00288 | .00400 | 0.472 |
| | constant | 4.32375 | .25006 | 0.000 |
| $\mu_2$ | *trt* | −.00138 | .00356 | 0.699 |
| | *pd* | −.00944 | .00319 | 0.003 |
| | *trt_pd* | −.00450 | .00329 | 0.171 |
| | constant | −.01899 | .00528 | 0.000 |
| logit($p$) | constant | −1.83391 | .26811 | 0.000 |
| $\ln(\alpha)$ | constant | −2.52851 | .49048 | 0.000 |

Initial health status for each component depends on baseline covariates representing previous treatments *prev*, level of beta-2 microglobulin *lgb2*, and baseline age *age*. The mean parameter for each component depends on treatment *trt*, whether PD occurred *pd*, and their interaction *trt-pd*.

benefit of this treatment regimen is enhanced somewhat if the patient is seen to experience disease progression (i.e., if *pd* equals 1). The fact that the mean parameter is positive for VELCADE has an interesting implication that we discuss later.

We now turn to the survival function of the second component, which is $F_2(r)$ in (6). This component is the majority component of the mixture. We see that the initial health level is adversely affected by the indicator for previous treatment *prev* and by the baseline level of beta-2 microglobulin *lgb2*. The large negative impact on survival of the latter covariate is evident with the regression coefficient being −.46291 for $\ln(x_{02})$. This effect represents about a 37% drop in initial health status for each unit increase in *lgb2* because $\exp(-.46291) = .63$. The treatment regimen has no effect on the mean parameter $\mu_2$ for this second survival component, although survival is affected by whether disease progression is observed or not The regression coefficient for *pd* is negative with borderline $P$-value 0.003, indicating that patients slip more quickly toward death when disease progression is observed while the patients are on their primary medication.

The estimated effects of the treatment regimen are summarized in Table 2, where the estimated mean parameters $\mu_1$ and $\mu_2$ for the two survival components of the mixture distribution are summarized for all combinations of *trt* and *pd*. The table shows the mean estimates for $\mu_1$ are positive for both VELCADE regimens. The positive values imply a positive probability that death will not occur from multiple myeloma with VELCADE as the primary treatment. The "cure"

**Table 2** Estimated mean parameters $\mu_1$ and $\mu_2$ for the two survival components of the mixture distribution for all four treatment regimens

| | | | | | Est. means | |
|---|---|---|---|---|---|---|
| Primary therapy | *trt* | *pd* | *trt_pd* | Patients | $\mu_1$ | $\mu_2$ |
| Dexamethasone | −1 | −1 | 1 | 107 | −.3203 | −.0127 |
| Dexamethasone | −1 | 1 | −1 | 193 | −.0692 | −.0226 |
| VELCADE | 1 | −1 | −1 | 160 | .0214 | −.0064 |
| VELCADE | 1 | 1 | 1 | 143 | .0745 | −.0343 |

probabilities are not large—varying from .024 to .120, depending on characteristics of the patients. Young patients with only one previous treatment ($prev = 0$) and low levels of beta-2 microglobulin, who are on VELCADE until PD, have the highest chance of experiencing nonfatal multiple myeloma. As multiple myeloma has been seen to date as a uniformly fatal condition, these small probabilities imply that VELCADE offers these patients a glimmer of hope for long life.

We now consider the composite time parameter $\alpha$, which is the rate of PD prior to the therapeutic switching event *pd*. We note that the mixture model with composite time reduces to a mixture model based on calendar time if $\alpha = 1$. We can see in Table 1, however, that this parameter differs significantly from 1. The parameter $\alpha$ enters the regression model with a logarithmic link function. The estimate of $\ln(\alpha)$ is significantly negative. The point estimate of $\alpha$ is $\exp(-2.52851) = 0.080$. The fact that it is substantially less than 1 signals that health status declines slowly until the switch to the alternate therapy. This result seems sensible as the switch to the alternate therapy will occur only when the patient's PD accelerates significantly under the primary therapy (of either kind). The regression analysis reported in Table 1 was redone with the treatment indicator variable added to the regression function for the composite time coefficient $\alpha$. The resulting regression coefficient was insignificant (*P*-value 0.787).

## COMPARISON TO COX REGRESSION

The TR regression results were compared to Cox proportional hazards regression results to check on the extent of agreement of the methods. Of course, exact agreement is impossible because the TR model based on the inverse Gaussian distribution is not a proportional hazards model. In addition, the TR regression model, and especially the mixture model defined on composite time, has a more intricate mathematical structure than the proportional hazards model. Table 3 shows the Cox regression output for this data set. The Cox regression coefficients show broad agreement with the TR results in Table 1 with *prev* and *lgb2* significantly magnifying the baseline hazard function of the proportional hazards model. The treatment variable *trt* favors VELCADE because the coefficient is significant (*P*-value = 0.028) and negative (−0.14484). The regression coefficient indicates that VELCADE shrinks the baseline hazard function by $\exp[-2(0.14484)] = 0.7485$ relative to Dexamethasone. The occurrence of PD while on the primary therapy ($pd = 1$) adds significantly to the hazard level.

**Table 3** Cox proportional hazards regression results for the study

| Variable | Estimate | Std. Error | *P*-value |
|---|---|---|---|
| *prev* | .41368 | .13679 | 0.002 |
| *lgb2* | .69033 | .09692 | 0.000 |
| *age* | .00297 | .00645 | 0.645 |
| *trt* | −.14484 | .06609 | 0.028 |
| *pd* | .20062 | .06592 | 0.002 |
| *trt_pd* | .07623 | .06572 | 0.246 |

The interaction term *trt_pd* is not significant. The Cox regression results are affected little if the interaction term is dropped from the model. Although the Cox regression results agree broadly with the TR results, TR provides more subtle insights into the source and nature of the comparative benefits of VELCADE than offered by the Cox methodology.

## ACKNOWLEDGMENTS

This research is supported in part by NIOSH Grant OH008649 (Lee) and by NSERC Grant 4032-05 (Whitmore). The authors thank Millennium Pharmaceutical Inc. for making background information on multiple myeloma, the study protocol and the data set available to the authors and for its financial support of the research.

## REFERENCES

Aalen, O. O., Gjessing, H. K. (2001). Understanding the shape of the hazard rate: A process point of view. *Statistical Science* 16:1–22.

Aalen, O. O., Gjessing, H. K. (2004). Survival models based on the Ornstein–Uhlenbeck process. *Lifetime Data Analysis* 10:407–423.

Balka, J. (2005). Wiener Process Cure Rate Models. Ph.D. Thesis, Guelph, Canada: University of Guelph.

Branson, M., Whitehead, W. (2002). Estimating a treatment effect in survival studies in which patients switch treatment. *Statistics in Medicine* 21:2449–2463.

Chhikara, R. S., Folks, J. L. (1989). *The Inverse Gaussian Distribution: Theory, Methodology and Applications*. New York: Marcel Dekker.

Cox, D. R., Miller, H. D. (1965). *The Theory of Stochastic Processes*. London: Chapman and Hall.

Duchesne, T., Lawless, J. F. (2000). Alternative time scales and failure time models. *Lifetime Data Analysis* 6:157–179.

Horrocks, J. C., Thompson, M. E. (2004). Modelling event times with multiple outcomes using the Wiener process with drift. *Lifetime Data Analysis* 10(10):29–49.

Lawless, J. F. (2003). *Statistical Models and Methods for Lifetime Data*. 2nd ed. Hoboken, NJ: Wiley.

Lee, M.-L. T., Whitmore, G. A. (2003). First hitting time models for lifetime data. In: Rao, C. R., Balakrishnan, N., eds. *Handbook of Statistics: Volume 23, Advances in Survival Analysis*. North Holland: Elsevier, pp. 537–543.

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

Lee, M.-L. T., DeGruttola, V., Schoenfeld, D. (2000). A model for markers and latent health status. *J. Royal. Statist. Soc., Series B* 62:747–762.

Oakes, D. (1995). Multiple time scales in survival analysis. *Lifetime Data Analysis* 1:7–18.

O'Brien, P. C., Fleming, T. R. (1979). Multiple testing procedure for clinical trials. *Biometrics* 35:549–556.

Padgett, W. J., Tomlinson, M. A. (2004). Inference from accelerated degradation and failure data based on Gaussian process models. *Lifetime Data Analysis* 10:191–206.

Padgett, W. J., Tomlinson, M. A. (2005). Accelerated degradation models for failure based on geometric Brownian motion and gamma processes. *Lifetime Data Analysis* 11:511–527.

Royston, P., Parmar, M. K. B. (2002). Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects. *Statistics in Medicine* 21:2175–2197.

Shao, J., Chang, M., Chow, S. C. (2005). Statistical inference for cancer trials with treatment switching. *Statistics in Medicine* 24:1783–1790.

Whitmore, G. A., Su, Y. (2007). Modeling low birth weights using threshold regression: results for U.S. birth data. *Lifetime Data Analysis* 13:161–190.