Piecewise Cox Models With Right-Censored Data

By George Y. C. Wong,[a] Michael P. Osborne[b], Qinggang Diao[c] and Qiqing Yu[c]

a Department of Integrative Medicine, Mount Sinai Beth Israel, New York, NY 10003
drgeorgewong@yahoo.com

b Strang Cancer Prevention Institute, New York, NY 10022

and

c Department of Mathematical Sciences, SUNY, Binghamton, NY 13902

Corresponding author's email address: qyu@math.binghamton.edu **AMS 1991 subject classification:** Primary 62 G07; Secondary 62 J99.

Short Title: Piecewise Cox Models.

Summary: We study a general class of piecewise Cox models. We discuss the computation of the semi-parametric maximum likelihood estimates (SMLE) of the parameters, with right-censored data, and a simplified algorithm for the maximum partial likelihood estimates (MPLE). Our simulation study suggests that the relative efficiency of the PMLE of the parameter to the SMLE ranges from 96% to 99.9%, but the relative efficiency of the existing estimators of the baseline survival function to the SMLE ranges from 3% to 24%. Thus the SMLE is much better than the existing estinators.

Key words and phrases: Cox model, semi-parametric MLE, time-dependent covariates.

1. Introduction. Time-dependent proportional hazards (PH) models involve a covariate vector $z(t)$ that is time dependent. Let Y be a random survival time variable that is subject to right censoring. Denote its survival distribution and density function by $S(t)$ and $f(t)$, respectively. The hazard function of Y is $h(t) = f(t)/S(t-)$. A time-dependent PH (TDPH) model for Y is represented by

$$h(t|z(t)) = h_o(t)e^{\varrho z(t)}, \tag{1}$$

where h_o is the baseline hazard function, ϱ is the vector of regression coefficients, $z(t)$ is the time-dependent covariate vector and $\varrho z(t)$ is the inner product of ϱ and $z(t)$. If Y is not continuous, then we need to add the restriction $t < \sup\{t : h_o(t) > 0\}$. Yu (2006) gives an example of a discrete h_o without the restriction such that (1) does not define a hazard function. We shall only consider the case that Y is continuous in this paper.

TDPH models have been discussed for right-censored (RC) data (see, for examples, Therneau and Grambsch (2000), Zhou (2001), Platt *et al.* (2004), Zhang and Huang (2006), Stephan and Michael (2007), Masaaki and Masato. (2009), and Leffondre *et al.* (2010)). A common situation that will involve the use of a TDPH model is a long-term clinical follow-up study. In such a study, the impact of a prognostic variable may change at different time periods. We have collected survival and relapse data on 371 women with early stage breast cancer with a median follow-up time of 7.4 years. The primary objective of the study is to investigate whether bone marrow micrometastasis (BMM) is significant in predicting early or late relapse. Clinical consideration and survival plots suggest early failure can be considered at time less than 4 years from initial breast cancer surgery. Our research interest is to employ a TDPH model to study the prognostic significance of BMM (presence or absence), together with standard clinical variables including lymph node involvement LN (yes or no) and tumor diameter TD ($>$ 2cm or $\leq$ 2cm). To allow for the possibility of differential impact of BMM for follow-up time ≤ 4 years and that > 4 years in a TDPH model, we let $z_1(t) = BMM \cdot \mathbf{1}(t \leq 4)$

and $z_2(t) = BMM \cdot \mathbf{1}(t > 4)$, where $\mathbf{1}(\cdot)$ denotes the indicator function. If BMM is the only variable in the model, then (1) yields

$$h(t|BMM) = \begin{cases} h_o(t)e^{\alpha BMM} & t \leq 4 \\ h_o(t)e^{\beta BMM} & t > 4. \end{cases} \tag{2}$$

The TDPH model (2) is an example of a piecewise PH (PWPH) model with one variable involving a single cut point at $t = 4$ years. PWPH models refer to regression models that are PH when restricted to different time intervals. Note that $z_{2i}(\mathbf{1}(t \leq c_i), \mathbf{1}(t > c_i))$ can be viewed as an (external) time-dependent covariate vector (see Kalbfleisch and Prentice (1980) p.123), or $\beta(t) = \beta_1 \mathbf{1}(t \leq c_i) + \beta_2 \mathbf{1}(t > c_i)$ can be viewed as a time varying regression coefficient (see Tian (2005)).

Zhou (2001) considered a more general PWPH model as follows.

$$h(t|\mathbf{z}_i) = h_o(t) \exp(\alpha z_{1i} + \beta_1 z_{2i} \mathbf{1}(t \leq c_i) + \beta_2 z_{2i} \mathbf{1}(t > c_i)), \tag{3}$$

where z_{ji} is a time-independent covariates (for $j \in \{1, 2, 3\}$), and c_i is the cut point for the i-th observation. The maximum partial likelihood estimator (MPLE) of the parameter can be obtained from an algorithm proposed by (Therneau (1999)). The method consists of first expanding the original data set and then apply a Cox regression-type algorithm to the resulting expanded data set. When $c_i = c$ for all i, we say the design of a PWPH model is balanced. Our research are concerned with balanced PWPH models.

In Section 2, we present the general form of a PWPH model with one cut point. We also show how to re-parameterize the model to facilitate derivation. We introduce a re-parametrization of the model in order to facilitate derivation and discussion. In Section 3, we demonstrate that the MPLE of a general class of PWPH models can be obtained in a piecewise fashion by appealing to a standard statistical program for time-independent PH (TIPH) models. Our algorithm is simpler than the Therneau algorithm for two reasons. First, we deal with the original unexpanded data set. Second, our algorithm optimizes over

smaller component parameter space, while the Therneau algorithm optimizes over the entire composite parameter space. Thus it is less likely to have a singular Hessian matrix. Data sets that our method works but not the Therneau method are given in Appendix (see Examples A1 and A2). We derive the generalized likelihood function for PWPH models in Section 4. To obtain the semiparametric maximum likelihood estimator (SMLE) of the model parameters by a numerical method, it is important to have a good initial value for the parameter vector. For this purpose, we propose a simple initial estimator and show that it is consistent. To assess the validity of the piecewise PH assumption, we propose in Section 5 a simple diagnostic method based on the SMLE of the survival function. We call this plot the PWPH plot. This method can also help us identify an appropriate cut point. The typical diagnostic method in the literature is the log-minus-log plots (see, for instance, Kalbfleisch and Prentice (1980)), and one checks whether the curves are parallel. Several other diagnostic plotting methods based on the scaled Schoenfeld residuals or the martingale residuals have also been proposed in the literature (see, for example, Therneau and Grambsch (2000)), but these methods needs to specify the forms of the covariate $z(t)$ (see Eq. (1)). Using our PWPH plot, one just needs to check whether the plot is piecewise linear and the "corner" of the piecewise linear graph suggests the cutting point, without specifying the form of $z(t)$ in advance. In Section 6, we discuss extension of our results to PWPH models involving more than one cut point. Simulation results are presented in Section 7. Finally in Section 8, we present survival analysis on our long-term breast cancer follow-up data using the results for the PWPH models we have developed. It is interesting to point out that it is not clear whether our data fit the PWPH models from the log-minus-log plots, but it is clear from the PWPH plots we proposed in this paper.

2. Model formulation. A general PWPH model with one cut point c should accommodate a covariate vector w that is time independent (same impact both to the left side and to the

right side of c), a set of covariates x_1 that follows a PH model to the left side of c but is absent in the PH model to the right side of c, a set of covariates x_2 that follows a PH model to the right side of c but is absent in the PH model to the left side of c, and a set of covariates z that follows a different PH model to the left side and to the right side of c. We represent the model as follows:

$$h(t|w, x_1, x_2, z) = \begin{cases} h_o(t)\exp(\gamma w + \rho_1 x_1 + \alpha z) & t \le c \\ h_o(t)\exp(\gamma w + \rho_2 x_2 + \beta z) & t > c, \end{cases} \tag{4}$$

where γ, ρ_1, ρ_2, α and β are vectors of regression coefficients.

To facilitate presentation and calculation, we can re-express (4) in a more compact formulation as follows.

$$h(t|w, u, v) = h_o(t)\exp(\gamma w + \alpha u \mathbf{1}(t \le c) + \beta v \mathbf{1}(t > c)), \tag{5}$$

where w, u and v are covariate vectors and it is possible that $u = v$. The time-dependent covariate vector of the general PWPH model is

$$z(t) = (w, u\mathbf{1}(t \le c), v\mathbf{1}(t > c)), \tag{6}$$

with corresponding independent parameter vector

$$\eta = (\gamma, \alpha, \beta). \tag{7}$$

Note that x_1 in (4) is represented by the coordinate covariates u_k (of u) such that $u_k \neq v_j$ for any j, x_2 by v_k such that $v_k \neq u_j$ for any j, and z by the coordinate covariates of u and v such that $u_k = v_k$.

3. Partial likelihood estimation. Let $M_i = \min(Y_i, C_i)$, where Y_i is the i-th survival time and C_i is the i-th censoring time, and δ_i be the indicator function for $M_i = Y_i$. The RC data for a TDPH model (1) (or PWPH model (5)) consist of $(M_i, \delta_i, z_i(t))$, $i = 1, ..., n$. For a TIPH model $h(t|z) = \exp(\beta z)h_o(t)$, the partial likelihood is given by

$$l_o = \prod_{i \in \mathcal{D}} \frac{\exp(\varrho z_i)}{\sum_{k \in \mathcal{R}_i} \exp(\varrho z_k)}, \tag{8}$$

where $\mathcal{R}_i$ is the risk set $\{j : \ M_j \geq Y_i\}$ and $\mathcal{D}$ is the collection of indices of the exact observations (see Cox and Oak (1984)). For a TDPH model (1), the partial likelihood is given by

$$l_1 = \prod_{i \in \mathcal{D}} \frac{\exp(\varrho z_i(M_i))}{\sum_{k \in \mathcal{R}_i} \exp(\varrho z_k(M_i))}. \tag{9}$$

Under the PWPH model (5) with observations $(M_i, \delta_i, w_i, u_i, v_i)$ and cut point c, the partial likelihood (9) becomes

$$l_1 = \prod_{i \in \mathcal{D}} \frac{exp(\gamma w_i + \alpha u_i \mathbf{1}(M_i \leq c) + \beta v_i \mathbf{1}(M_i > c))}{\sum_{k \in \mathcal{R}_i} exp(\gamma w_k + \alpha u_k \mathbf{1}(M_i \leq c) + \beta v_k \mathbf{1}(M_i > c))}. \tag{10}$$

When there is no w in (5), the hazard function for (5) is given by

$$h(t|u,v) = \begin{cases} h_o(t)e^{\alpha u} & t \leq c \\ h_o(t)e^{\beta v} & t > c. \end{cases}$$

The last expression leads to a simpler algorithm which is stated in Theorem 1 that says that the MPLE $\hat{\alpha}_P$ of α can be computed with a standard statistical program for a TIPH model by censoring data M_i's at c, and for the MPLE $\hat{\beta}_P$ of β using data $M_i > c$.

Theorem 1. *Let $w = 0$ in (5). The maximum partial likelihood estimate $(\hat{\alpha}_P, \hat{\beta}_P)$ of (α, β) can be obtained as follows:*

(I) For computing $\hat{\alpha}_P$, modify the data by right censoring all M_i's with $M_i > c$ at c. Let

$$(M_i^*, \delta_i^*, u_i^*) = \begin{cases} (M_i, \delta_i, u_i) & \text{if } M_i \leq c \\ (c, 0, u_i) & \text{if } M_i > c, \end{cases} \tag{11}$$

and use an existing statistical software for the TIPH model $h(t|u^*) = h_o(t)\exp(\alpha u^*)$.

(II) For computing $\hat{\beta}_P$, use only the data (M_i, δ_i, v_i) satisfying $M_i > c$ and use an existing statistical software for the TIPH model $h(t|v) = h_o(t)\exp(\beta v)$.

Proof. Without loss of generality (WLOG), we can assume that the data are ordered in the M_i's, and $M_m \leq c < M_{m+1}$. The partial likelihood (9) becomes

$$\begin{aligned} l_1 &= \prod_{i \in \mathcal{D}} \frac{exp(\alpha u_i \mathbf{1}(M_i \leq c) + \beta v_i \mathbf{1}(M_i > c))}{\sum_{k \in \mathcal{R}_i} exp(\alpha u_k \mathbf{1}(M_i \leq c) + \beta v_k \mathbf{1}(M_i > c))} \\ &= \prod_{i \leq m} [\frac{exp(\alpha u_i)}{\sum_{k \geq i} exp(\alpha u_k)}]^{\delta_i} \prod_{i > m} [\frac{exp(\beta v_i)}{\sum_{k \geq i} exp(\beta v_k)}]^{\delta_i}. \end{aligned} \tag{12}$$

In l_1, here are two parameters α and β. The partial likelihood l_1 is a product of two factors. The first factor does not depend on β and the second factor does not depend on α. Thus we only need to maximize the first and the second factors separately to obtain $\hat{\alpha}_P$ and $\hat{\beta}_P$, respectively.

The first factor in (12) is the same as the partial likelihood for the modified data $(M_i^*, \delta_i^*, u_i^*)$ defined in (11). Therefore, a standard statistical package for the TIPH model can be used here to obtain $\hat{\alpha}_P$. The message of (11) is that to compute the MPLE of the model parameters to the left side of the cut point c, we need only use data $M_i \leq c$ and treat all data $M_i > c$ as censored at c.

The second factor in (12) is the same as the partial likelihood in (8) for the truncated data removing all the $M_i \leq c$. Such a truncated data set satisfies the conditional distribution

$$\frac{S(t|z)}{S(c|z)} = \frac{(S_o(t))^{e^{\beta v}}}{(S_o(c))^{e^{\beta v}}} = (S_1(t))^{e^{\beta v}},$$

where $S_o(t) = S(t|0)$ is the baseline survival function and $S_1(t) = \frac{S_o(t)}{S_o(c)}$. Therefore, the truncated data set satisfies a TIPH model. Then the partial likelihood solution of β from a standard package is the same as $\hat{\beta}_P$ based on the entire data set. □

We now describe the application of Therneau's algorithm (1999) to fit the MPLE associated with the model $h(w_i, u_i, v_i) = h_o(t)e^{\gamma w_i + \alpha u_i \mathbf{1}(t \leq c_i) + \beta v_i \mathbf{1}(t > c_i)}$. The program is written in Splus, R or SAS. The command in R is coxph(Surv(start,stop,event)∼z). The original data set is $(M_i, \delta_i, w_i, u_i, v_i)$ with cut point c_i, $i = 1, ..., n$. Let m be the number of $M_i \leq c_i$. The data set is expanded as follows.

Step 1. For each $M_i \leq c_i$ generate one vector

$$(start_{i1}, stop_{i1}, \delta_{i1}^*, w_{i1}^*, u_{i1}^*, v_{i1}^*) = (0, M_i, \delta_i, w_i, u_i, 0).$$

Step 2. For each $M_i > c_i$, generate two vectors

$$\begin{cases} (start_{i1}, stop_{i1}, \delta_{i1}^*, w_{i1}^*, u_{i1}^*, v_{i1}^*) = (0, c, 0, w_i, u_i, 0) \\ (start_{i2}, stop_{i2}, \delta_{i2}^*, w_{i2}^*, u_{i2}^*, v_{i2}^*) = (c, M_i, \delta_i, w_i, 0, v_i). \end{cases}$$

Let *start*, *stop*, δ^*, w^*, u^*, and v^* be the $(2n-m)$ dimensional vectors with coordinates $start_{ij}$, $stop_{ij}$, δ^*_{ij}, w^*_{ij}, u^*_{ij}, and v^*_{ij}, respectively. Then the MPLE $(\hat{\gamma}, \hat{\alpha}, \hat{\beta})$ can be obtained by applying the command coxph(Surv($start, stop, \delta^*$) $\sim w^* + u^* + v^*$).

When we apply our simple algorithm and the Therneau algorithm to our BMM data where $c_i = c$ for all i, both of the algorithms yield the same MPLE values for almost all data under all the models studied there. In the Appendix, we present two data examples (see Examples A1 and A2) that our algorithm provides the solutions but Therneau algorithm does not because the 2×2 Hessian matrix is singular for these two data sets. Thus our algorithm has the advantage over Therneau algorithm in simplicity and feasibility.

4. Full Likelihood Estimation. For TIPH models, Finkelstein (1986) proposes the full likelihood approach by making use of the generalized likelihood of Kiefer and Wolfowitz (1956):

$$\mathcal{L} = \prod_{i=1}^{n} ((S_o(M_i))^{e^{\beta z}})^{1-\delta_i} ((S_o(M_i-))^{e^{\beta z}} - (S_o(M_i))^{e^{\beta z}})^{\delta_i},$$

where S_o is the baseline survival function. For TDPH model (1) with RC data (M_i, δ_i, z_i),

$$\mathcal{L} = \prod_{i=1}^{n} [(S(M_i|z_i))^{1-\delta_i} (S(M_i - |z_i) - S(M_i|z_i))^{\delta_i}]. \tag{13}$$

By model assumption in Eq (5), (6) and (7), $S(\cdot|\cdot)$ is a function of S_o and the parameter η. The SMLE of (S_o, η) maximizes $\mathcal{L}$ over all η and over all S_o in the collection

$$\{S: \ S \text{ is non-increasing, right continuous and } S \in [0,1]\}$$

Evaluation of the function $\mathcal{L}$ in (13) requires an explicit expression for $S(t|z)$ in terms of S_o and the paramter. We derive this for the general continuous PWPH model (5) in Theorem 2.

Theorem 2. If Y is continuous, the survival function corresponding to model (5) is given by

$$S(t|w,u,v) = \begin{cases} (S_1(t))^{e^{\alpha u}} & t \le c \\ (S_1(c))^{e^{\alpha u}} (\frac{S_1(t)}{S_1(c)})^{e^{\beta v}} & t > c, \end{cases} \quad \text{where } S_1(t) = (S_o(t))^{e^{\gamma w}}. \tag{14}$$

Proof. Since Y is continuous, $S(t|w,u,v) = exp(-\int_0^t h(x|w,u,v)dx)$. When $t \le c$,

$$\begin{aligned} S(t|w,u,v) =& exp(-\int_0^t e^{\gamma w+\alpha u} h_o(x)dx) \\ =& (S_o(t))^{e^{\gamma w+\alpha u}} \\ =& (S_1(t))^{e^{\alpha u}}. \end{aligned}$$

When $t > c$,

$$\begin{aligned} S(t|w,u,v) =& \exp(-\int_0^c e^{\gamma w+\alpha u} h_o(x)dx) \exp(-\int_c^t e^{\gamma w+\beta v} h_o(x)dx) \\ =& (S_o(c))^{e^{\gamma w+\alpha u}} (\frac{S_o(t)}{S_o(c)})^{e^{\gamma w+\beta v}} \\ =& (S_1(c))^{e^{\alpha u}} (\frac{S_1(t)}{S_1(c)})^{e^{\beta v}}. \ \square \end{aligned}$$

Remark 1. It is well known (see Wong and Yu (2012)) that under the non-parametric set-up with RC data, the SMLE of the TIPH model only assigns weights to exact observations or ∞. It can be shown that in order to maximize $\mathcal{L}$ under the PWPH model, in addition to put probability weights of S_o to exact observations or ∞, one may also need to put the probability weight of S_o to c.

Let t_1, ..., t_N be the distinct exact observations of Y_i or c or ∞, where

$$-\infty = t_0 < t_1 < \cdots < t_m = c < t_{m+1} < \cdots < t_N = \infty.$$

For each i, let $r_i, l_i \in \{0, 1, ..., N\}$ such that $(t_{r_i}, t_{l_i}) = \begin{cases} (M_i, t_{r_i-1}) & \text{if } \delta_i = 1 \\ (t_N, \max\{t_j : \ t_j \le M_i\}) & \text{if } \delta_i = 0. \end{cases}$
Substituting the expression for $S(M_i|z_i) = S(M_i|w_i,u_i,v_i)$ in Theorem 2 into $\mathcal{L}$ in (13), we obtain the full likelihood for model (5).

$$\begin{aligned} \mathcal{L} =& \prod_{i:\ M_i \le c,\ \delta_i=0} (S_1(M_i))^{e^{\alpha u_i}} \prod_{i:\ M_i > c,\ \delta_i=0} (S_1(c))^{e^{\alpha u_i - \beta v_i}} (S_1(M_i))^{e^{\beta v_i}} \\ & \times \prod_{i:\ M_i \le c, \delta_i=1} ((S_1(M_i-))^{e^{\alpha u_i}} - (S_1(M_i))^{e^{\alpha u_i}}) \end{aligned}$$

$$\times \prod_{i:\ M_i>t_{m+1},\delta_i=1} (S_1(c))^{e^{\alpha u_i}-e^{\beta v_i}}((S_1(M_i-))^{e^{\beta v_i}}-(S_1(M_i))^{e^{\beta v_i}})$$
$$\times \prod_{i:\ M_i=t_{m+1},\delta_i=1} ((S_1(c))^{e^{\alpha u_i}}-(S_1(c))^{e^{\alpha u_i}-e^{\beta v_i}}(S_1(M_i))^{e^{\beta v_i}}),$$

where $S_1(M_i) = (S_o(t))^{e^{\gamma w_i}}$, $S_1(c)^{\alpha u_i+\beta v_i} = (S_o(t))^{e^{\gamma w_i+\alpha u_i+\beta v_i}}$, and an arbitrary S_o puts weights to the $2N$ disjoint intervals (t_0,t_1), $[t_1,t_1]$, ..., (t_{N-1},t_N), $[t_N,t_N]$. Notice that for each $j \in \{1,...,N\}$, $\mathcal{L}$ increases if the positive weight assigned to the interval (t_{j-1},t_j) is moved to t_j. Thus in order to obtain the SMLE of $S(\cdot|\cdot)$, it suffices to obtain the SMLE $(\hat{\eta},\hat{S}_o)$ of (η,S_o) which maximizes the likelihood

$$\begin{aligned}\mathcal{L} =& g(S_o(c)) \prod_{M_i\in[t_1,c)} (S_o(t_{l_i}))^{e^{(\gamma_i w_i+\alpha u_i)}(1-\delta_i)} \prod_{M_i>c} (S_o(t_{l_i}))^{e^{(\gamma w_i+\beta v_i)}(1-\delta_i)} \qquad (15)\\ &\times \prod_{M_i<c} ((S_o(t_{l_i}))^{e^{(\gamma w_i+\alpha u_i)}}-(S_o(M_i))^{e^{(\gamma w_i+\alpha u_i)}})^{\delta_i}\\ &\times \prod_{M_i>t_{m+1}} ((S_o(t_{l_i}))^{e^{(\gamma w_i+\beta v_i)}}-(S_o(M_i))^{e^{(\gamma w_i+\beta v_i)}})^{\delta_i},\\ &\text{subject to the constraint } 0\le S_o(t_1)\le\cdots\le S_o(t_{N-1})\le S_o(t_N)=1,\end{aligned}$$

where $g(S_o(c)) = (S_o(c))^{\sum_{i:\ M_i\ge t_{m+1}} e^{\gamma w_i+\alpha u_i}-e^{\gamma w_i+\beta v_i}+\sum_{t_{l_i}=c}(1-\delta_i)e^{\gamma w_i+\alpha u_i}}$

$$\times \prod_{M_i=t_{m+1}} ((S_o(c))^{e^{\gamma w_i+\beta v_i}}-(S_o(t_{m+1}))^{e^{\gamma w_i+\beta v_i}})^{\delta_i}$$
$$\times \prod_{M_i=c} ((S_o(t_{l_i}))^{e^{(\gamma w_i+\alpha u_i)}}-(S_o(c))^{e^{(\gamma w_i+\alpha u_i)}})^{\delta_i}.$$

The SMLE has to be obtained by an iterative algorithm. It is well known (see Wong and Yu (2012)) that in general the Newton-Raphson method does not work for the SMLE under the Cox regression model, we implement a steep decent method to compute the SMLE.

Remark 2. For each $j \in \{1,...,N-1\}$, given $S_o(t_i)$ for $i \notin \{j,N\}$, in general, there is no closed form solution for the maximum point of $\mathcal{L}$ with respect to $S_o(t_j)$, except for $j=m$, *i.e.* $t_j = c$. In the latter case, given $S_o(t_i)$ for $i \notin \{m,N\}$, the maximum point of $\mathcal{L}$ with respect to $S_o(t_m)$ $(= S_o(c))$ has an explicit solution, given in Lemma 1. Notice that $S_o(c)$ only occurs in $\mathcal{L}$ through $g(S_o(c))$ (see Eq. (15)). Thus it suffices to maximizes $g(x)$.

Let $g(x) = (x^y - t)x^z$ (as $\sum_i \mathbf{1}(M_i = t_{m+1}) = 1$ and $\sum_{i=1}^n \mathbf{1}(M_i = c, \delta_i = 1) = 0$ by the continuity assumption on Y), where (y, t, z)

$$= (e^{\gamma w_{i_1} + \beta v_{i_1}}, (S_o(t_{m+1}))^{e^{\gamma w_{i_1} + \beta v_{i_1}}}, \sum_{i:\ M_i \geq t_{m+1}} (e^{\gamma w_i + \alpha u_i} - e^{\gamma w_i + \beta v_i}) + \sum_{i: t_{l_i} = c} (1 - \delta_i) e^{\gamma w_i + \alpha u_i}),$$

and i_1 satisfies that $M_{i_1} = t_{m+1}$ and $\delta_{i_1} = 1$.

Lemma 1. *$g(x) \leq g(x_o^{\xi_o}(S_o(t_{m-1}))^{1-\xi_o})$ for $x \in [S_o(t_{m+1}), S_o(t_{m-1})]$, where $x_o = (\frac{tz}{y+z})^{1/y}$ and $\xi_o = \mathbf{1}(x_o \in (S_o(t_{m+1}), S_o(t_{m-1}))$.*

Proof. It is easy to verify the following statements:

1. $(\log g(x))' = \frac{z}{x} + \frac{x^y y/x}{x^y - t} = (z + \frac{x^y y}{x^y - t})/x = 0$ yields the unique root $x_o = (\frac{tz}{y+z})^{1/y}$;
2. both $g(x)$ and $(\log g(x))'$ exist if $x^y > t$, that is, $x \in [S_o(t_{m+1}), \infty)$;
3. $(\log g(x))'$ change its sign only at x_o;
4. $g(S_o(t_{m+1})) = 0$;
5. $g(x) > 0$ if $x > S_o(t_{m+1})$, that is, $x^y > t$ (≥ 0).

If $x_o \notin (S_o(t_{m+1}), S_o(t_{m-1}))$, then $(\log g(x))'$ does not change sign in the same interval due to statements 3. Thus it is increasing in x in the interval $[S_o(t_{m+1}), S_o(t_{m-1})]$ due to statements 4 and 5. Consequently, $g(x)$ achieves its maximum at $x = S_o(t_{m-1})$ if $x \in [S_o(t_{m+1}), S_o(t_{m-1})]$.

If $x_o \in (S_o(t_{m+1}), S_o(t_{m-1}))$, $g(x)$ achieves its extremum uniquely at the stationary point x_o due to statements 1, 2 and 3. If $g(x_o)$ is the minimum point, then $g(x_o) < 0$ by statement 4, but it contradicts statement 5. Thus $g(x)$ achieves its maximum at x_o. □

In general, the likelihood function (15) is not a convex function. If one implements a numerical algorithm to find $(\hat{\eta}, \hat{S}_o)$, it is important to use a good initial value η_0 for η to increase the chance of convergence and the chance of convergence to the correct solution. We propose a simple algorithm to derive a consistent estimate of η and use it as η_0. The main idea is to replace the time independent component γw with $[\gamma \mathbf{1}(t \leq c) + \gamma^* \mathbf{1}(t > c)]w$, and apply the results of Theorem 1. The algorithm goes as follows.

1. Let $(M_i^*, \delta_i^*, w_i^*, u_i^*) = \begin{cases} (M_i, \delta_i, w_i, u_i) & \text{if } M_i \le c \\ (c, 0, w_i, u_i) & \text{if } M_i > c. \end{cases}$ Use a standard package to fit a TIPH model to the data $(M_i^*, \delta_i^*, w_i^*, u_i^*)$'s and obtain estimate $(\tilde{\gamma}, \tilde{\alpha})$ for (γ, α) for $M_i \le c$.
2. Fit a TIPH model to the data $(M_i, \delta_i, w_i, v_i)$'s satisfying $M_i > c$ and obtain estimate $(\tilde{\gamma}^*, \tilde{\beta})$ for (γ, β).
3. Set $\tilde{\eta} = (\tilde{\gamma}, \tilde{\alpha}, \tilde{\beta})$ or $(\tilde{\gamma}^*, \tilde{\alpha}, \tilde{\beta})$ and let $\eta_0 = \tilde{\eta}$.

We now explain why $\tilde{\eta}$ is consistent. When we regard w as piecewise at c, the PWPH model becomes

$$h(t|w_i, u_i, v_i) = \begin{cases} h_o(t)\exp(\gamma w_i + \alpha u_i) & \text{if } t \le c \\ h_o(t)\exp(\gamma^* w_i + \beta v_i) & \text{if } t > c. \end{cases}$$

The likelihood corresponding to this hazard function is the same as the partial likelihood (12) if we identify u_i in (12) with (w_i, u_i) and v_i with (w_i, v_i), α with (γ, α) and β with (γ, β). By Theorem 1, $(\tilde{\gamma}, \tilde{\alpha}) = \hat{\alpha}_P$ and $(\tilde{\gamma}^*, \tilde{\beta}) = \hat{\beta}_P$ are consistent, as it is well known that the MPLE is consistent. Note that both $\tilde{\gamma} \to \gamma$ and $\tilde{\gamma}^* \to \gamma$ with probability 1.

5. Diagnostic plots. We propose a simple diagnostic method for PWPH models to assess any departure from the PWPH assumption and the appropriateness of the cut point. We present our diagnostic method for a PWPH model with a simple dichotomous variable that impacts differently on the left and right sides of a cut point c. In expression(14), set $w \equiv 0$ and $u = v$. Then

$$S(t|u) = \begin{cases} (S_o(t))^{e^{\alpha u}} & \text{if } t \le c \\ (S_o(c))^{e^{\alpha u}} (\frac{S_o(t)}{S_o(c)})^{e^{\beta u}} & \text{if } t > c. \end{cases}$$

Let u take values 0 and 1. Then

$$S(t|0) = S_o(t), \ \forall\ t; \tag{16}$$

$$S(t|1) = \begin{cases} (S_o(t))^{e^{\alpha}} & \text{if } t \le c \\ (S_o(c))^{e^{\alpha}} (\frac{S_o(t)}{S_o(c)})^{e^{\beta}} & \text{if } t > c. \end{cases} \tag{17}$$

Equivalently,

$$\ln S(t|1) = \begin{cases} e^{\alpha} \ln S_o(t) & \text{if } t \le c \\ e^{\beta} \ln S_o(t) + (e^{\alpha} - e^{\beta}) \ln S_o(c) & \text{if } t > c. \end{cases} \tag{18}$$

We propose a diagnostic plotting procedure as follows.

1. Obtain Kaplan-Meier estimates (KME) $\hat{S}_o$ of S_o by calculating the KME $\hat{S}(t|0)$ (see (16)) based on observations (M_i, δ_i)'s with $u_i = 0$, and obtain the KME $\hat{S}(t|1)$ of $S(t|1)$ (see (17)) based on the observations (M_i, δ_i)'s with $u_i = 1$.
2. Plot $-ln\hat{S}(t|1)$ on the y-axis against $-ln\hat{S}_o(t)$ on the x-axis (see (18)).
3. If the plot consists of one linear line segment going through the origin, then it suggests that the PWPH model is simply a TIPH model. If the plot consists of two linear line segments, then it suggests the PWPH model with one cut point (see middle right panel of Figure 1 in Section 8). If the plot shows K (> 1) linear line segments, then it suggests that a model with K cut points is appropriate (see Section 6 and the bottom right panel of Figure 1 in Section 8).

We remark that expression (18) suggests that the diagnostic plot can be used to estimate an appropriate value for c.

6. Extension to models involving multiple cut points. For ease of exposition and WLOG, we consider a general PWPH model with two cut points c_1 and c_2

$$h(t|u,r,v) = \begin{cases} h_o(t)e^{\gamma w+\alpha u} & \text{if } t \le c_1 \\ h_o(t)e^{\gamma w+\theta r} & \text{if } t \in (c_1, c_2] \\ h_o(t)e^{\gamma w+\beta v} & \text{if } t > c_2, \end{cases}$$

where r denotes a covariate vector that is present in $(c_1, c_2]$, and θ denotes the corresponding parameter vector. When $w = 0$, Theorem 1 can be extended in a straightforward manner as follows:

(1) For computing $\hat{\alpha}_P$, modify the data by right censoring all times M_i's at c_1. Let

$$(M_i^*, \delta_i^*, u_i^*) = \begin{cases} (M_i, \delta_i, u_i) & \text{if } M_i \le c_1 \\ (c_1, 0, u_i) & \text{if } M_i > c_1, \end{cases}$$

and use an existing statistical software for the TIPH model $h(t|u^*) = h_o(t)\exp(\beta u^*)$.

(2) For computing $\hat{\theta}_P$, use only the data satisfying $M_i > c_1$ and right censored them at c_2. Let $(M_i^*, \delta_i^*, r_i^*) = \begin{cases} (M_i, \delta_i, r_i) & \text{if } M_i \in (c_1, c_2] \\ (c_2, 0, r_i) & \text{if } M_i > c_2, \end{cases}$

and use an existing statistical software for the TIPH model $h(t|r^*) = h_o(t)\exp(\theta r^*)$.

(3) For computing $\hat{\beta}_P$, use only the data satisfying $M_i > c_2$ and use an existing statistical software for the TIPH model $h(t|v) = h_o(t)\exp(\beta v)$.

The full likelihood can be obtained by maximizing the likelihood

$$\mathcal{L}(\eta, S_o) = \prod_{i=1}^{n}(S(M_i - |w_i, u_i, r_i, v_i) - S(M_i|w_i, u_i, r_i, v_i))^{\delta_i}(S(M_i|w_i, u_i, r_i, v_i))^{1-\delta_i},$$

where $\eta = (\gamma, \alpha, \theta, \beta, S_o)$ and $S_o(c_j) = S_o(c_j-)$ if c_j is not an exact observation, $j = 1, 2$,

$$S(t|w, u, r, v) = \begin{cases} (S_1(t))^{e^{\alpha u}} & t \le c_1 \\ (S_1(c_1))^{e^{\alpha u}}(\frac{S_1(t)}{S_1(c_1)})^{e^{\theta r}} & t \in (c_1, c_2] \\ (S_1(c_1))^{e^{\alpha u}}(\frac{S_1(c_2)}{S_1(c_1)})^{e^{\theta r}}(\frac{S_1(t)}{S_1(c_2)})^{e^{\beta v}} & t > c_2, \end{cases}$$

where $S_1(t) = (S_o(t))^{e^{\gamma w}}$.

As for diagnostic plots, it follows directly from (17) with $w = 0$ that the extension of Expression (18) is given by

$$\ln S(t|1) = \begin{cases} e^{\alpha}\ln S_o(t) & \text{if } t \le c_1 \\ e^{\theta}\ln S_o(t) + (e^{\alpha} - e^{\theta})\ln S_o(c_1) & \text{if } t \in (c_1, c_2] \\ e^{\beta}\ln S_o(t) + (e^{\alpha} - e^{\theta})\ln S_o(c_1) + (e^{\theta} - e^{\beta})\ln S_o(c_2) & \text{if } t > c_2. \end{cases} \tag{19}$$

Again the diagnostic method parallels that for the case of one cut point. A PWPH model with a dichotomous variable will exhibit three linear line segments. It will then be possible to estimate the values of c_1 and c_2 from the diagnostic plots, as is in the case of one cut point.

The Therneau algorithm for finding the MPLE of model (19) is as follows.

Recall that the data are $(M_i, \delta_i, w_i, u_i, r_i, v_i)$, $i = 1, ..., n$, satisfying model

$$h(t|u_i, r_i, v_i) = h_o(t)\exp(\gamma w_i + \alpha u_i \mathbf{1}(t \in (c_0, c_1]) + \theta r_i \mathbf{1}(t \in (c_1, c_2]) + \beta v_i \mathbf{1}(t \in (c_2, c_3]))$$

where $c_0 = 0 < c_1 < c_2 < c_3 = \infty$ are the cut points.

Step 1. For each $M_i \le c_1$ generate one vector

$(start_{i1}, stop_{i1}, \delta^*_{i1}, w^*_{i1}, u^*_{i1}, r^*_{i1}, v^*_{i1}) = (0, M_i, \delta_i, w_i, u_i, 0, 0)$.

Step 2. For each $M_i \in (c_1, c_2]$, generate two vectors

$$\begin{cases} (start_{i1}, stop_{i1}, \delta^*_{i1}, w^*_{i1}, u^*_{i1}, r^*_{i1}, v^*_{i1}) = (0, c_1, 0, w_i, u_i, 0, 0) \\ (start_{i2}, stop_{i2}, \delta^*_{i2}, w^*_{i2}, u^*_{i2}, r^*_{i2}, v^*_{i2}) = (c_1, M_i, \delta_i, w_i, 0, r_i, 0). \end{cases}$$

Let n_2 be the number of $M_i \in (c_1, c_2]$.

Step 3. For each $M_i > c_2$, generate 3 vectors

$$\begin{cases} (start_{i1}, stop_{i1}, \delta^*_{i1}, w^*_{i1}, u^*_{i1}, r^*_{i1}, v^*_{i1}) = (0, c_1, 0, w_i, u_i, 0, 0) \\ (start_{i2}, stop_{i2}, \delta^*_{i2}, w^*_{i2}, u^*_{i2}, r^*_{i2}, v^*_{i2}) = (c_1, c_2, 0, w_i, 0, r_i, 0) \\ (start_{i3}, stop_{i3}, \delta^*_{i3}, w^*_{i3}, u^*_{i3}, r^*_{i3}, v^*_{i3}) = (c_2, M_i, \delta_i, w_i, 0, 0, v_i). \end{cases}$$

Let n_3 be the number of $M_i > c_2$.

Let $start$, $stop$, δ^*, w^*, u^*, r^* and v^* be the $(n+n_2+2n_3)$ dimensional vectors with coordinates $start_{ij}$, $stop_{ij}$, δ^*_{ij}, w^*_{ij}, u^*_{ij}, r^*_{ij} and v^*_{ij}, respectively. Then the MPLE $(\hat{\gamma}, \hat{\alpha}, \hat{\theta}, \hat{\beta})$ can be obtained by applying the command

$$\text{coxph}(\text{Surv}(start, stop, \delta^*) \sim w^* + u^* + r^* + v^*).$$

7. Simulation Studies. We study the performance of the SMLE under several simulation assumptions. We drew data (Y_i, u_i)'s from the PWPH model $h(t|u_i) = e^{\alpha u_i 1(t \le c) + \beta u_i 1(t > c)} h_o(t)$, where Y_i is right censored at 3, u_i is from Binomial$(1, 0.5)$, the baseline survival function S_o is either from Exp(1) or from $U(0, 4)$, and c is either 0.15 or 1. In the simulation, we assume that α is a known value, so only β is unknown. We generated data with 5000 replications each for sample sizes $n = 100$, or 400, or 800. Table 1 displays the results.

Our simulation study suggests that the SMLE $\hat{\beta}$ is consistent and the convergence rate is $\sqrt{n}$. The SMLE $\hat{\beta}$ is just a little bit better than the PMLE $\hat{\beta}_P$.

We also compare the PMLE to the SMLE of parameters under the same assumptions. Our simulation study suggests that the relative efficiency of the PMLE of the parameter to the SMLE ranges from 96% to 99.9%.

Moreover, we compare the estimators of the baseline survival function S_o. Notice that the SMLE of S_o is well defined by the full likelihood in Eq. (15), denoted by $\hat{S}_o$, but there is no PMLE of S_o, even under the Cox model with the time-independent covariates, as the partial likelihood function does not involve S_o. Several estimators of S_o making use of the

PMLE of β were proposed in the literature. In Table 1, we only present the default one in the functions coxph() and survfit() of R program, denoted by $\tilde{S}_o$. It is seen from Table that the SMLE $\hat{S}_o$ is much better than $\tilde{S}_o$, whose relative efficiency ranges from 3% to 24%.

n	(α,β)	S_o	c	$\hat{\beta}$	$SD_{\hat{\beta}}$	$\hat{\beta}_P$	$SD_{\hat{\beta}_P}$	$\frac{SD_{\hat{\beta}}}{SD_{\hat{\beta}_P}}$
100	(0,1)	Exp(1)	1	1.047	0.411	1.044	0.415	0.990
400	(0,1)	Exp(1)	1	1.009	0.180	1.007	0.189	0.956
100	(-1,2)	Exp(1)	0.15	2.010	0.313	2.014	0.314	0.997
400	(-1,2)	Exp(1)	0.15	1.993	0.145	1.985	0.145	0.999
400	(0,-2)	U(0,4)	1	-1.932	0.248	-2.014	0.255	0.972
800	(0,-2)	U(0,4)	1	-1.932	0.167	-2.015	0.174	0.955
400	(-2,3)	U(0,4)	1	2.950	0.212	3.012	0.220	0.962
800	(-2,3)	U(0,4)	1	2.961	0.145	3.015	0.148	0.983

n	(α,β)	S_o	c	$\hat{S}_o(1.1)$	$SD_{\hat{S}_o(1.1)}$	$\tilde{S}_o(1.1)$	$SD_{\tilde{S}_o(1.1)}$	$\frac{SD_{\hat{S}_o(1.1)}}{SD_{\tilde{S}_o(1.1)}}$
100	(0,1)	Exp(1)	1	0.334	0.004	0.333	0.017	0.235
400	(0,1)	Exp(1)	1	0.333	0.001	0.333	0.008	0.125
100	(-1,2)	Exp(1)	0.15	0.331	0.004	0.339	0.062	0.065
400	(-1,2)	Exp(1)	0.15	0.333	0.001	0.334	0.031	0.032
			$S_o(1.1)$	0.333				
400	(0,-2)	U(0,4)	1	0.724	0.001	0.725	0.010	0.100
800	(0,-2)	U(0,4)	1	0.724	0.0006	0.725	0.007	0.086
400	(-2,3)	U(0,4)	1	0.726	0.0009	0.725	0.006	0.150
800	(-2,3)	U(0,4)	1	0.725	0.0005	0.725	0.004	0.125
			$S_o(1.1)$	0.725				

Table 1. Simulation results for SMLE and PMLE of β and S_o

8. Data analysis. Our data are obtained from an Institutional Review Board approved long-term clinical follow-up study on 371 women with stages I-III unilateral invasive breast cancer treated by surgery at Memorial Sloan-Kettering Cancer Center in New York City between 1985 and 2001.

The primary objective of the study is to investigate the prognostic significance of bone marrow micrometastasis (BMM) for relapse and survival. The censoring rate is 87%. The median follow-up time of the study is 7.4 years (range is 1 month-180 months (14.8 years)),

which is the longest among published studies on BMM. For the discussion here, we will consider only the covariates BMM, lymph node involvement LN and tumor diameter TD as mentioned in Section 1. Meta-analysis of BMM by Braun *et. al.* (2005) suggests that a stepwise Cox model with a cut point at 4 years is appropriate for the BMM data. We fitted the PWPH model (5) with $w \equiv 0$, $u = v = (BMM, LN, TD)$ and $c = 4$ years to the data. We obtained the MPLE and the SMLE of the model parameters. Table 2 summarizes the multivariate results.

MPLE	BMM	p-value	LN	p-value	TD	p-value
< 4 years	0.40	0.32	1.04	0.03	0.51	0.23
> 4 years	0.20	0.65	0.80	0.10	0.46	0.31
SMLE	BMM	p-value	LN	p-value	TD	p-value
< 4 years	0.30	0.22	1.24	0.00	0.38	0.19
> 4 years	0.31	0.21	0.62	0.06	0.62	0.07

Table 2. Multivariate analysis of BMM data

Multivariate analysis of the PWPH model by either the partial likelihood method or the full likelihood method indicates that BMM does not predict survival either before 4 years or after 4 years. Before four years, LN is the only variable that was significant both by partial likelihood analysis and by full likelihood analysis. The fact that the important clinical variable TD was not significant in the multivariate analysis can be explained by its correlation with LN. After four years, partial likelihood shows that LN has lost its prognostic significance and TD remains not predictive. However, full likelihood analysis suggests that both LN and TD may still be predictive. Our finding here is in direct contrast to the conclusions from all recent published studies (for example, see Braun *et al.* (2005)) that conclude that BMM involvement is a strong predictor of poor survival by TIPH multivariate analysis.

We applied both the log-minus-log plots and our PWPH plots to assess the appropriateness of the PWPH model with a single cut point at 4 years and to see if the PWPH model can be improved. The diagnostic plots for BMM, LN and TD are give in Figure 1.

It can be seen from Figure 1 that each pair of the standard log-minus-log plots does not appear parallel along the y-axis, except perhaps the top left panel in Figure 1. The top left plots suggest that the two curves appear somewhat coincide, thus it suggests that the data fit the Cox model with time-independent covariate BMM. It is not easy to see any patterns from the other two log-minus-log plots.

On the other hand, the PWPH plot for BMM at the top right panel of Figure 1 yields roughly a simple straight line. Therefore, our diagnostic plot for a dichotomous covariate suggests that BMM should follow a TIPH model. In the general PWPH model (5), BMM corresponds to the covariate w. The message here is that BMM does not differentially predict early and late failure. BMM was not significant by univariate TIPH regression using partial likelihood method.

The PWPH diagnostic plot for LN in the middle right panel of Figure 1 shows two distinct straight line segments. Our diagnostic method suggests that a PWPH model for LN with a single cut point is appropriate. We estimated the slope and the intercept (see (13)) of each of the two line segments by linear regression and concluded that a reasonable choice of c is a value between 4 and 5 years. When $c = 4$. univariate analysis by partial likelihood yields the MPLE $\hat{\alpha}_P = 1.24$ with $p = 0.005$ before 4 years and $\hat{\beta}_P = 0.96$ with $p = 0.03$ after 4 years.

The PWPH diagnostic plot for TD at the bottom right panel in Figure 1 displays three straight line segments. Our diagnostic method suggests a PWPH model for TD should have two cut points c_1 and c_2. From (14) and using linear regression, $c_1 = 1.5$ and $c_2 = 6$. Since the two-year and five-year marks are recognized time points of clinical relevance in cancer follow-up, we choose $c_1 = 2$ and $c_2 = 5$.

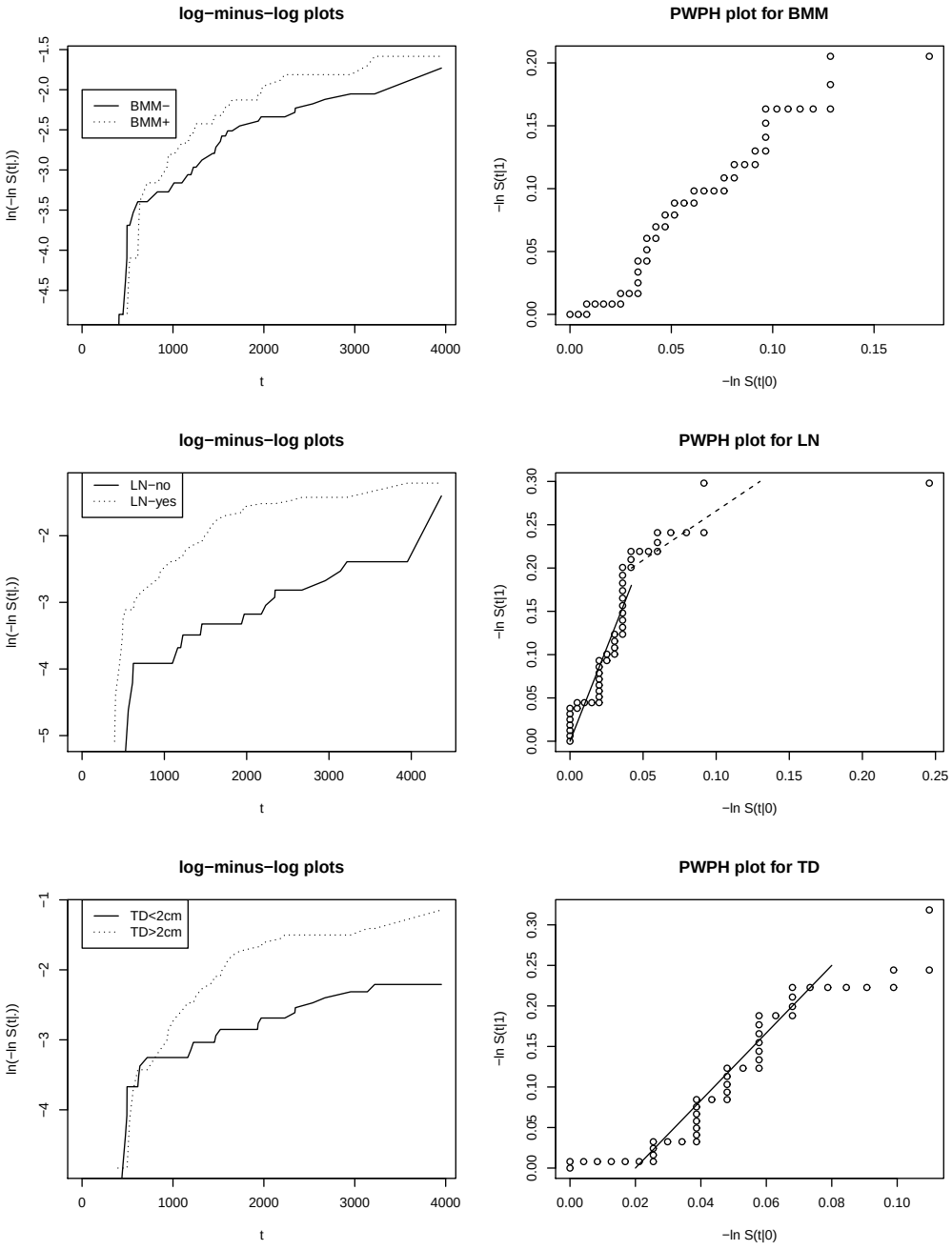

Figure 1. Diagnostic plots for BMM, LN and TD

We obtained the 3 MPLEs for TD with two cut points at 2 and 5 as -1.76, 2.02 and 0.34, with p-value 0.77, 0.001 and 0.52, respectively.

Our univariate analysis of TD by a PWPH model with two cut points suggests that TD

does not predict early deaths or late failure of patients who have achieved five-year survival. But TD is a very significant predictor between two and five years. The conclusions here differ from the conventional view that TD possess the same significant predictor power for survival throughout a breast cancer patient's life time.

The above analysis suggests that our long-term breast cancer follow-up data is best described by a PWPH model with a time-independent component for BMM, a piecewise component with a single cut point at $c = 4$ years for LN, and a piecewise component with two cut points at $c_1 = 2$ years and $c_2 = 5$ years. However, we will need more data and a longer follow-up time to yield a more informative diagnostic plot for each of the three covariates. This is particularly true for the diagnostic plot for TD in which more data information will enable us to gain a better assessment of the lower and upper time segments. We propose to update and add about 800 new cases to the BMM database.

The proposed re-analysis of the BMM data, however, will require a PWPH model involving unequal cut points (at 4 years for LN, and at 2 and 5 years for TD). Further research is needed to extend the results developed for PWPH models with equal cut points to those with unequal cut points. We shall again consider both the partial likelihood approach and the full likelihood approach. We expect that the full likelihood SMLE method should be able to be extended to the PWPH model with unequal cut points.

Acknowledgement. Funding for George Wong is provided by American Foundation For Chinese Medicine, and for Michael Osborne by Strang Cancer Prevention Institute. Qiqing Yu is partially supported by NSF Grant DMS- 1106432. The authors thank for the editor and a referee for valuable comments on the paper.

References.

* Braun, S., Vogl, F. D., Naume, B., Janni, W., Osborne, M.P., Coombes, R.C., Schlimok, G., Diel, I.J., Gerger, B., Gebauer, G., Pierga, J.Y., Marth, C., Oruzio, D., Wiedswang,

G., Solomayer, E.F., Kundt, G., Strobl, B., Fehm, T., Wong, G.Y.C., Bliss, J., Salomon, A.V. and Pantel, K. (2005). A pooled analysis of bone marrow micrometastasis in breast cancer. *New England Journal of Medicine*, 353 793-802.

* Cox, D.R. and Oakes, D. (1984). Analysis of Survival Data. *Chapman & Hall* NY.

* Finkelstein, D.M. (1986). A proportional hazards model for interval-censored failure time data. *Biometrics*, 42 845-854.

* Kalbfleisch, J.D. and Prentice, R. L. (1980). The statistical analysis of failure time data. *Wiley* NY.

* Kiefer, J and Wolfowitz, J. (1956). Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters. *Ann. Math. Statist.*, 27, 887-906.

* Masaaki, T. and Masato, S. (2009). Analysis of survival data having time-dependent covariates. *IEEE Transaction on Neural Networks*, 20 389-394.

* Leffondre K., Wynant W., Cao Z., et al. (2010). A weighted Cox model for modeling time-dependent exposures in the analysis of case-control studies. *Statistics in Medicine*, 29 839-850.

* Platt, R.W., Joseph3, K. S., Ananth, C. V., Grondines1, J., Abrahamowicz, M, and Kramer, M. S. (2004). A proportional hazards model with time-dependent covariates and time-varying effects for analysis of fetal and infant death. *American Journal of Epidemiology*, 160 199-206.

* Stephan, L. and Michael, S. (2007). Parsimonious analysis of time-dependent effects in the Cox model. *Statistics in Medicine*, 26 2686-2698.

* Therneau, T., (1999) A package for survival analysis in S.
http://www.mayo.edu/hsr/people/therneau/surviva.ps

* Therneau, T., and Grambsch, P. (2000). Modeling survival data : extending the Cox model. Springer.

* Tian, L., Zucker, D. dn Wei, L. J. (2005). On the Cox model with time-varying regression coefficients. *Journal of the American Statistical Association* 100 172-183.

* Yu, Q.Q. (2006). A note on the proportional hazards model with discrete data. *Statistics and Probability Letters*, 77 735-739.

* Wong, G.Y.C and Yu, Q.Q (2012). Estimation under the Lehmann regression model with interval-censored data. *Comm. Statist. — Comput. Simul.*, 41 1489-1500.

* Zhang H.M. and Huang C.F. (2006). Nonparametric survival analysis on time-dependent covariate effects in case-cohort sampling design. *Statistica Sinica*, 16 267-285.

* Zhou, M. (2001). Understanding the Cox regression models with time-change covariates. American Statistian, 55 153-155

Appendix.

Example A.1. Consider the model $h(t|z) = \exp((\alpha \mathbf{1}(t \leq 1) + \beta \mathbf{1}(t > 1))z)h_o(t)$. A data set (y, δ, z) of 50 RC data with the cut point $c = 1$ are given as follows.

$> y$

[1] 0.01889 0.03201 0.03507 0.10677 0.11106 0.23184 0.29353 0.40049 0.50323 2.65036

[11] 0.51341 0.54032 0.54246 0.56325 0.57291 0.69315 0.74532 0.83477 0.88320 2.69411

[21] 0.97981 1.00453 1.01508 1.03283 1.04153 1.04333 1.04565 1.04864 1.06527 2.92332

[31] 1.06746 1.07607 1.11337 1.12039 1.12285 1.12499 1.17092 1.21515 1.27879 3.04159

[41] 1.28950 1.44392 1.47707 1.47778 1.57985 1.60793 1.61542 1.87746 1.94818 4.46560

$> \delta$

[1] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0

[39] 0 0 1 1 0 0 1 0 0 0 0 1

$> z$

[1] 0 0 0 1 0 0 0 0 0 1 0 0 0 0 0 0 0 1 0 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1

[39] 0 0 1 1 0 0 0 0 0 0 0 0

Applying the method in Theorem 1 to R program yields $(\hat{\alpha}, \hat{\beta}) = (-1.837, 22.548)$, but the Therneau algorithm yields (NA, NA).

Example A.2. Consider the model $h(t|z) = exp((\alpha \mathbf{1}(t \leq 1) + \beta \mathbf{1}(t > 1))z)$. A data set of 100 RC data with the cut point $c = 1$ are given as follows. Applying the method in Theorem 1 to R program yields $(\hat{\alpha}, \hat{\beta}) = (-1.66, 20.52)$, but the Therneau algorithm presents $(NA, 32.58)$. Notice that the estimates of β using both algorithms indicate $\beta = \infty$.

$> y$

[1] 0.00495 0.02291 0.04044 0.06118 0.06879 0.06904 0.07210 0.08047 0.08462

[10] 0.08825 0.10049 0.10824 0.12244 0.14294 0.19162 0.21194 0.21499 0.24943

[19] 0.26049 0.26478 0.27132 0.29989 0.30673 0.34700 0.35673 0.39163 0.40032

[28] 0.41490 0.41914 0.43303 0.43604 0.51220 0.55450 0.60633 0.61418 0.64214

[37] 0.70096 0.72695 0.73642 0.79251 0.81413 0.84694 0.89843 0.93752 0.95366

[46] 0.97138 1.00119 1.00262 1.00339 1.00463 1.00505 1.00734 1.01196 1.01643

[55] 1.01964 1.02023 1.02238 1.02270 1.02362 1.02414 1.02627 1.02815 1.02907

[64] 1.03241 1.05252 1.05454 1.05543 1.05737 1.05928 1.06516 1.06763 1.07549

[73] 1.08590 1.08643 1.10780 1.10832 1.11567 1.11636 1.12387 1.14569 1.16403

[82] 1.17903 1.19876 1.20499 1.22961 1.26507 1.27199 1.29644 1.33341 1.34185

[91] 1.42591 1.49448 1.51977 1.63795 1.65927 1.71104 1.81075 2.29736 2.85665

[100] 4.38886

$> \delta$

[1] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

[38] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

[75] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0

$> z$

[1] 0 0 1 1 1 0 0 0 0 1 0 0 1 0 0 0 0 1 0 0 1 0 0 0 0 1 1 0 0 0 0 0 0 1 1 0 0

[38] 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

[75] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0