Published in final edited form as:

Biometrics. 2017 June ; 73(2): 463–472. doi:10.1111/biom.12580.

Semiparametric Time-to-Event Modeling in the Presence of a Latent Progression Event

John D. Rice* and Alex Tsodikov**

Department of Biostatistics, University of Michigan, Ann Arbor, Michigan, U.S.A

Summary

In cancer research, interest frequently centers on factors influencing a latent event that must precede a terminal event. In practice it is often impossible to observe the latent event precisely, making inference about this process difficult. To address this problem, we propose a joint model for the unobserved time to the latent and terminal events, with the two events linked by the baseline hazard. Covariates enter the model parametrically as linear combinations that multiply, respectively, the hazard for the latent event and the hazard for the terminal event conditional on the latent one. We derive the partial likelihood estimators for this problem assuming the latent event is observed, and propose a profile likelihood-based method for estimation when the latent event is unobserved. The baseline hazard in this case is estimated nonparametrically using the EM algorithm, which allows for closed-form Breslow-type estimators at each iteration, bringing improved computational efficiency and stability compared with maximizing the marginal likelihood directly. We present simulation studies to illustrate the finite-sample properties of the method; its use in practice is demonstrated in the analysis of a prostate cancer data set.

Keywords

Survival analysis; semiparametric methods; expectation-maximization algorithm

1. Introduction

The analysis of time-to-event data in areas of biomedical research dealing with progressive diseases is particularly challenging because a large part of disease progression is not observed. For example, a disease is typically diagnosed only when symptoms reach the point where a patient seeks medical attention, or is detected through some screening program, with the point of onset of detectable disease being unobserved. Another example studied in this paper is metastatic progression in prostate cancer, where cancer-specific death occurs due to metastasis, whose onset is unobserved. It is often the case that we are interested in the effect of covariates on time to onset of disease as well as time to diagnosis, so we are confronted with a combination of the usual right censoring characteristic of survival analysis

*jdrice@umich.edu

**tsodikov@umich.edu

Supplementary Materials

Web Appendices referenced in Sections 2–6 are available with this paper at the *Biometrics* website on Wiley Online Library (code used for fitting the models is provided as Appendix I).

(Kalbfleisch and Prentice, 2002) as well as left- or more complex censoring as in our particular application (Dejardin et al., 2010; Tsodikov et al., 1995).

To make these concepts concrete, let T_1 denote the time to the terminal event, and let T_0 denote the time to the latent event. We assume that T_1 is observed, but is subject to right censoring, which is indicated by $\delta = 0$; $\delta = 1$ means that the terminal event occurs at time T_1 . The time T_0 , by contrast, is never observed, therefore generally we must rely on T_1 and the structure of the model to inform us about the distribution of T_0 .

Our work draws on the literature of frailty models, which is vast, dating back to Vaupel et al. (1979), who introduced the concept of frailty variables in life-table analysis and assumed a gamma distribution. An important difference between our model and much of the previous work on frailty models, however, is that our frailty is not a random variable but rather a stochastic process $N_0(t)$ that jumps from 0 to 1 at the time of the latent event, and reflects the fact that the subject is not at risk of the terminal event until the latent event has already occurred. While there has been some work done on such models (e.g., Gjessing, Aalen, and Hjort, 2003), the overwhelming body of research considers only frailties that are properly random variables, that is, are fixed at time 0. Hu and Tsodikov (2014b) develop a similar model for cancer progression that also makes use of a jump process as the frailty. However, they did not devise an efficient EM approach, which is the key contribution of the present paper. Also, theirs is a marked survival response model where the latent event does not necessarily precede the terminal event.

Sternberg and Satten (1999) propose a model for chains of events which relies on an assumption of independence between interevent times, although they are dealing with interval censored data that results from only examining subjects periodically, and are in the discrete-time setting. Likewise, Frydman (1995), working in a discrete-time framework, proposes a model that involves nonparametric estimation of time to some intermediate event and semiparametric estimation of time to a terminal event; again, however, the intermediate event times are observed subject to interval censoring. Frydman and Szarek (2009) also propose a multi-state Markov model and derive nonparametric maximum likelihood estimators, but in their scenario there is no natural ordering to the two events: one is assumed to be nonfatal but related to the disease process, while the other is death. Lin, Sun, and Ying (1999) also deal with events that have a recurrent ordering in time with the goal of jointly modeling the gap time distribution between serial events, the primary statistical problem there being dependent censoring induced by the time ordering of the events.

The multi-state model proposed by Dejardin et al. (2010) with two events, progression (of cancer) and death, is similar to ours in that it also assumes a (recurrent) ordering to the two events, with progression necessarily preceding death. However, they specify a flexible parametric form for the baseline hazard, while we propose an EM algorithm (Dempster et al., 1977) to estimate it nonparametrically.

That the shape of the hazard function is shared by both the latent and terminal events is the underlying assumption behind our model. It is motivated by the idea that there is a common process driving both the latent and terminal events, as would be the case for example with

the growth of a tumor: its growth pattern will determine both the time to onset of detectable disease (latent event) and time to diagnosis (terminal event). The interpretation of the common baseline hazard function in this case would be as a surrogate for the tumor-growth process. Inference for the parametric part of the model is based on standard profile likelihood theory for semiparametric models (Murphy and van der Vaart, 2000; Tsodikov, 2003).

The structure of this paper is as follows: in Section 2, we formally introduce the model and give the associated likelihood function; in Section 3, we describe the general procedure for deriving nonparametric maximum likelihood estimators and use it to derive the EM algorithm for our model; Section 4 gives the results of a simulation study to examine the performance of the model on its own terms (bias, agreement between estimated and true standard errors, etc.) as well as compared to a parametric alternative when the model is misspecified; Section 5 applies our method to a SEER prostate cancer data set; finally, a discussion is given in Section 6.

2. Model and likelihood

2.1 Data structure and notation

There are two events associated with our model, latent and terminal. The time to the latent event is denoted as T_0 and is never observed; time to the terminal event is T_1 . By definition, the latent event must precede the terminal one: $T_0 \leq T_1$. There is a censoring time C that is (conditional on covariates) independent of T_0 and T_1 . We observe $(T^*, \delta, \mathbf{z}')$, where $T^* = \min\{T_1, C\}$ and $\delta = \mathbb{1}(T^* = T_1)$ and $\mathbf{z}$ is a vector of covariates; $\mathbb{1}(\cdot)$ is the indicator function, taking the value 1 if $\cdot$ is true and 0 otherwise. The maximum follow-up time is τ .

Note that if $\delta = 1$, we must have $T_0 \leq T_1 \leq C$. However, if $\delta = 0$, then either $C \leq T_0$ or $T_0 < C \leq T_1$. Thus, we are unable to tell from observed data whether or not the latent event has occurred in the case of a censored observation.

2.2 Model

As in Dejardin et al. (2010), we formulate our model in two parts. The first is the marginal hazard of the latent event $d\Lambda_0$, and the second is the conditional hazard of the terminal event given time to the latent event $d\Lambda_1$.

$$d\Lambda_0(t|\mathbf{z}) = \lim_{h \rightarrow 0} \frac{P(T_0 \in [t, t+h) | T_0 \geq t, \mathbf{z})}{h} = \mu dH(t) \quad (1)$$

$$d\Lambda_1(t|T_0, \mathbf{z}) = \lim_{h \rightarrow 0} \frac{P(T_1 \in [t, t+h) | T_1 \geq t, T_0, \mathbf{z})}{h} = (t > T_0) \eta dH(t). \quad (2)$$

The latent and terminal events may be thought of as two events in a recurrent-events model: when $\eta = \mu$ for a subject, these will be two events in a Poisson process. For $\eta > \mu$, the terminal event will be accelerated following the latent event (relative to such a Poisson process), while for $\eta < \mu$ the reverse is true. The baseline hazard $H(\cdot)$ models the temporal pattern of the disease progression (see Hu and Tsodikov, 2014b, for a mechanistic justification and detailed discussion). Covariates $\mathbf{z}$ will enter the model through μ and η :

specifically, for $\beta = (\beta_0, \beta'_\eta, \beta'_\mu)'$, we have $\eta = \eta(\beta) = e^{\beta_0 + \mathbf{z}'\beta_\eta}$ and $\mu = \mu(\beta) = e^{\mathbf{z}'\beta_\mu}$. For notational simplicity we refer to a single covariate vector $\mathbf{z}$ and assume that it contains all covariates relevant to the model, but it would be possible to restrict some components of either β_η or β_μ to be zero.

Because the multiplicative effect in (2) is a random process $\mathbb{1}(t > T_0) = 1 - Y_0(t)$, where $Y_0(t) = \mathbb{1}(T_0 \leq t)$ is an unobserved at risk process for the intermediate event, the model belongs to a class of dynamic stochastic process frailty models, where the distribution of the frailty process and the conditional model have a common infinite-dimensional parameter. No EM solutions are available for this class of problems to our knowledge. As ours is a full-likelihood approach, it is asymptotically fully efficient, in contrast to Breslow estimators based on martingale estimating equations.

2.3 Marginal distribution of time to the terminal event

Generally, when the conditional hazard function for a survival time T is a stochastic process,

say $u(t)$, the marginal distribution has survival function $G(t) = \left[e^{-\int_0^t u(x) dx} \right]$, density

function $g(t) = \left[u(t) e^{-\int_0^t u(x) dx} \right]$, and hazard function

$\lambda(t) = [u(t) | T > t] = g(t) / G(t) = \left[u(t) e^{-\int_0^t u(x) dx} \right] / \left[e^{-\int_0^t u(x) dx} \right]$, where the expectations are taken over the trajectory (t) of the random process u from 0 to t (see Gjessing, Aalen, and Hjort, 2003). For our specific model, the marginal survival function for the terminal event is the expectation of (7) over the distribution of T_0 for a censored observation:

$$G_*(t) = \frac{1}{\eta - \mu} \left[\eta e^{-\mu H(t)} - \mu e^{-\eta H(t)} \right]. \quad (3)$$

The marginal density function is given by the expectation of (7) for a failed observation:

$$g_*(t) = \eta \mu \frac{e^{-\mu H(t)} - e^{-\eta H(t)}}{\eta - \mu} dH(t). \quad (4)$$

The marginal hazard function is $\lambda_*(t) = g_*(t) / G_*(t)$. (See Supplementary Materials Appendix A for details.)

It may be observed that if μ and η are exchanged, (3) and (4) are unaltered, and an identifiability problem emerges. When the model contains the same set of covariates in both μ and η , for any given set of parameters $(\beta_0, \beta'_\eta, \beta'_\mu)'$, we have the same marginal distribution of time to the observed event (integrating over the unobserved time to the latent event) with $(-\beta_0, \beta'_\mu, \beta'_\eta)'$. The source of the issue is the fact that summands are exchangeable while the sum is fixed.

2.4 Conditional distribution of time to the latent event

The model also allows us to make predictions of the distribution of the time to the latent event, given observed data on a subject. This is of particular importance to clinical practice, as it allows us to examine, for a subject who has not experienced the terminal event after some specified time T^* , the distribution of the time to the latent event. Specifically, for the survival functions, we have (see Supplementary Materials Appendix B for details)

$$G(t_0|T^*) = \begin{cases} \frac{\eta e^{-\mu H(T^*)} - \mu e^{-\eta H(T^*) + (\eta - \mu)H(t_0)}}{\eta e^{-\mu H(T^*)} - \mu e^{-\eta H(T^*)}}, & t_0 < T^* \\ \frac{(\eta - \mu)e^{-\mu H(t_0)}}{\eta e^{-\mu H(T^*)} - \mu e^{-\eta H(T^*)}}, & t_0 \geq T^*, \end{cases} \quad (5)$$

for a censored subject, and

$$G(t_0|T^*) = \begin{cases} \frac{e^{(\eta - \mu)H(T^*)} - e^{(\eta - \mu)H(t_0)}}{e^{(\eta - \mu)H(T^*)} - 1}, & t_0 < T^* \\ 0, & t_0 \geq T^*. \end{cases} \quad (6)$$

for a failed subject (that is, a subject who has experienced the terminal event).

2.5 Likelihood

The likelihood for a single subject with observed data $(T^*, \cdot)$ conditional on time to the latent event $T_0 = t_0$ is

$$L_0 = [(t_0 < T^*)\eta dH(T^*)]^\Delta e^{-(t_0 < T^*)\eta[H(T^*) - H(t_0)]}. \quad (7)$$

Using (3) and (4), we can write the marginal log-likelihood associated with this model in counting process form:

$$\ell(H(t); \beta) = \sum_{i=1}^n \int_0^T \{[\log \gamma_i(H(t); \beta) + \log dH(t)] dN_i(t) - Y_i(t) \gamma_i(H(t); \beta) dH(t)\}, \quad (8)$$

where

$$\gamma_i(H(t); \boldsymbol{\beta}) = \eta_i \mu_i \frac{e^{-\mu_i H(t)} - e^{-\eta_i H(t)}}{\eta_i e^{-\mu_i H(t)} - \mu_i e^{-\eta_i H(t)}}$$

$$\text{and } \mu_i = e^{\mathbf{z}_i' \boldsymbol{\beta}_\mu}, \eta_i = e^{\beta_0 + \mathbf{z}_i' \boldsymbol{\beta}_\eta}.$$

We define the martingale $dM_i(t)$ based on observed counting processes $N_i(t)$ with respect to filtration $\mathcal{F}(t-) = \sigma\{N_i(s), Y_i(s), \mathbf{z}_i : s \in [0, t], i = 1, \dots, n\}$ as

$$dM_i(t) = dN_i(t) - Y_i(t) \gamma_i(H(t); \boldsymbol{\beta}) dH(t),$$

since

$$\begin{aligned} [dN_i(t) | \mathcal{F}_i(t-)] &= Y_i(t) P[dN_i(t) = 1 | Y_i(t) = 1] \\ &= Y_i(t) \gamma_i(H(t); \boldsymbol{\beta}) dH(t) \end{aligned}$$

by definition of the hazard rate under the true model. Note that $Y_i(t) = (T_i^* \geq t)$ is the at-risk process for subject i .

3. Nonparametric maximum likelihood estimation

3.1 Functional derivative and score equations

Define derivatives of γ_i with respect to H and $\boldsymbol{\beta}$ as

$$\dot{\gamma}_{i,H}(H(t); \boldsymbol{\beta}) = \frac{\partial \gamma_i(H(t); \boldsymbol{\beta})}{\partial dH(t)} \quad (9)$$

$$\dot{\gamma}_{i,\boldsymbol{\beta}}(H(t); \boldsymbol{\beta}) = \frac{\partial \gamma_i(H(t); \boldsymbol{\beta})}{\partial \boldsymbol{\beta}}, \quad (10)$$

respectively. For a functional $J(f)$, $f = f(x)$, the functional derivative in (9) is defined as

$$\frac{\partial J(f)}{\partial df(s)} = \frac{\partial J(f + \varepsilon g)}{\partial \varepsilon} \Big|_{\varepsilon=0, g=(x>s)}$$

(see Hu and Tsodikov, 2014a, Section 3.2) and corresponds to taking the derivative with respect to a jump in H at time t when H is a step function. For a linear functional of the form $J(f) = \int_0^t \varphi(x) df(x)$, the functional derivative is

$$\begin{aligned}\frac{\partial J}{\partial f(s)} &= \int_0^t \varphi(x) d \frac{\partial(f+\varepsilon g)}{\partial \varepsilon} \Big|_{\varepsilon=0, g=(x>s)} \\ &= \int_0^t \varphi(x) d(x>s) \\ &= \varphi(s) (t \geq s).\end{aligned}$$

Using this definition, we have

$$\begin{aligned}\frac{\frac{\partial H(t)}{\partial dH(s)}}{\frac{\partial \log dH(t)}{\partial dH(s)}} &= (t \geq s) \\ &= \frac{1}{dH(t)} \frac{\partial dH(t)}{\partial dH(s)} = \frac{1}{dH(t)} d(t \geq s).\end{aligned}$$

First we differentiate the log-likelihood with respect to the infinite-dimensional parameter $H(\cdot)$, making use of the identity $f(t) = f(t) \log f(t)$ to write in terms of the martingale $dM_i(t)$:

$$\mathcal{U}_{dH(s)} = \sum_{i=1}^n \left[\int_s^\tau \frac{\dot{\gamma}_{i,H}(H(t); \boldsymbol{\beta})}{\gamma_i(H(t); \boldsymbol{\beta})} dM_i(t) + \frac{dM_i(s)}{dH(s)} \right]. \quad (11)$$

In order to facilitate the asymptotic analysis, we replace s by a dummy variable x and integrate this expression to obtain the alternative form of the score:

$$\mathcal{U}_{H(s)} = \sum_{i=1}^n \int_0^\tau \left[\frac{\dot{\gamma}_{i,H}(H(t); \boldsymbol{\beta})}{\gamma_i(H(t); \boldsymbol{\beta})} H(s \wedge t) + (t < s) \right] dM_i(t). \quad (12)$$

Define $\varepsilon_i(t, s; H, \boldsymbol{\beta}) = \frac{\dot{\gamma}_{i,H}(H(t); \boldsymbol{\beta})}{\gamma_i(H(t); \boldsymbol{\beta})} H(s \wedge t) + (t < s)$. As shown in Hu and Tsodikov (2014a, Supplementary Materials B), the linear transform $\int_0^\tau \varepsilon_i(t, s; H, \boldsymbol{\beta}) dM_i(t)$ is a martingale as a process in s under the true model when $\varepsilon_i(t, s; H, \boldsymbol{\beta})$ does not depend on s for $t < s$, as is the case here. The score function for the regression parameters $\boldsymbol{\beta}$ is

$$\mathcal{U}_{\boldsymbol{\beta}} = \sum_{i=1}^n \int_0^\tau \frac{\dot{\gamma}_{i,\boldsymbol{\beta}}(H(t); \boldsymbol{\beta})}{\gamma_i(H(t); \boldsymbol{\beta})} dM_i(t). \quad (13)$$

3.2 Complete data profile (partial) likelihood

The complete data likelihood contribution for the i th subject is $L_{0i}P_{0i}$, where L_{0i} is defined as in equation (7) and P_{0i} is the likelihood of the latent event. In counting process notation, we have for the contribution of the i th subject to the log-likelihood

$$\log(L_{0i}P_{0i}) = \int_0^\tau \left\{ \log[\eta_i dH(t)] dN_i(t) - \tilde{Y}_i(t) \eta_i dH(t) \right\} + \int_0^\tau \left\{ \log[\mu_i dH(t)] dN_{0i}(t) - Y_{0i}(t) \mu_i dH(t) \right\}. \quad (14)$$

The counting processes $N_i(t)$ and $N_{0i}(t)$ count occurrences of the terminal and latent events, respectively; likewise, $\tilde{Y}_i(t) = 1(T_{0i} \leq t < T_i^*)$ is the at-risk process for the terminal event and $Y_{0i}(t) = 1(T_{0i} \leq t)$ is the at-risk process for the latent event.

It is straightforward to derive the NPMLE of the hazard now: we have

$$\widehat{dH}(s) = \frac{\sum_{i=1}^n [dN_i(s) + dN_{0i}(s)]}{\sum_{i=1}^n [\eta_i \tilde{Y}_i(s) + \mu_i Y_{0i}(s)]}; \quad (15)$$

details are given in Supplementary Materials Appendix C. Substitution of this estimator into (14) leads to the log profile likelihood

$$\begin{aligned} \ell_p(\boldsymbol{\beta}) = & \sum_{i=1}^n \int_0^\tau \{ \log \eta_i dN_i(t) + \log \mu_i dN_{0i}(t) \} \\ & - \sum_{i=1}^n \int_0^\tau \log \left(\sum_{j=1}^n [\eta_j \tilde{Y}_j(t) + \mu_j Y_{0j}(t)] \right) \\ & \times [dN_i(t) + dN_{0i}(t)]. \end{aligned} \quad (16)$$

We will call it a partial likelihood following the Cox model analogy, and to distinguish it from the profile likelihood for the marginal model where T_0 is unobserved.

3.3 EM algorithm

In the situation where the latent event is not observed for any subjects, and in view of the simplicity of the complete-data problem, it is natural to make use of the EM algorithm. Using the relative expectation approach of Tsodikov (2003), we have derived the EM algorithm for this problem; details are given in Supplementary Materials Appendix D. This results in the following self-consistency equation that defines iterations over $k = 0, 1, 2, \dots$, that converges $dH^{(k)} \rightarrow \widehat{dH}$ as $k \rightarrow \infty$.

$$\sum_{i=1}^n \left[\frac{dN_i(s)}{dH^{(k+1)}(s)} - \Psi_i^{(k)}(s) \right] = \sum_{i=1}^n \left[1 - \frac{dH^{(k)}(s)}{dH^{(k+1)}(s)} \right] \theta_i^{(k)}(s), \quad (17)$$

where

$$\Psi_i^{(k)}(s) = Y_i(s) \frac{\eta^{1-\Delta_i} \mu e^{-\mu H^{(k)}(T_i^*)} - \eta \mu^{1-\Delta_i} e^{-\eta H^{(k)}(T_i^*)}}{\eta^{1-\Delta_i} e^{-\mu H^{(k)}(T_i^*)} - \mu^{1-\Delta_i} e^{-\eta H^{(k)}(T_i^*)}}$$

and

$$\theta_i^{(k)}(s) = \frac{(\eta - \mu) \mu^{1-\Delta_i}}{\eta^{1-\Delta_i} e^{-\mu H^{(k)}(T_i^*)} - \mu^{1-\Delta_i} e^{-\eta H^{(k)}(T_i^*)}} \times \left\{ Y_i(s) e^{-\eta H^{(k)}(T_i^*) + (\eta - \mu) H^{(k)}(s)} + (1 - \Delta_i) [1 - Y_i(s)] e^{-\mu H^{(k)}(s)} \right\}.$$

Note that at the solution, i.e., when $dH^{(k)} = dH^{(k+1)} = \widehat{dH}$, the right-hand side of (17) vanishes, leaving the score equation for the marginal likelihood of the time to the terminal event.

Solving (17) for the next-iteration hazard, we obtain a Breslow-type expression

$$dH^{(k+1)}(s) = \frac{\sum_{i=1}^n dN_i(s) + \left[\sum_{i=1}^n \theta_i^{(k)}(s) \right] dH^{(k)}(s)}{\sum_{i=1}^n \left[\Psi_i^{(k)}(s) + \theta_i^{(k)}(s) \right]}. \quad (18)$$

The second term in the numerator may be thought of as the imputed $\sum_{i=1}^n dN_{0i}(s)$, while the denominator could be called an effective imputed at-risk process for the combined latent and terminal failures in the denominator of (15).

3.4 Estimation procedure

Now that we have described the EM algorithm used to estimate the baseline hazard for this model, we proceed with the general procedure for estimation.

1. Set $j = 0$. $\hat{\beta}^{(j)}$ is initialized with coefficient estimates from a Cox model fit.
2. Find $\hat{\beta}^{(j+1)}$ by taking one step toward maximizing $\ell(\hat{H}(\beta)(t); \beta)$ with respect to β using a conventional optimization routine (e.g., BFGS).
3. Find $\widehat{dH}^{(j+1)}(t) = d\hat{H}(\beta^{(j+1)})(t)$. Increment j .
4. Repeat steps (2)–(3) until

$$\ell\left(\hat{H}^{(j)}(t); \hat{\beta}^{(j)}\right) - \ell\left(\hat{H}^{(j-1)}(t); \hat{\beta}^{(j-1)}\right) < \varepsilon$$

The implicit function $\hat{H}(\beta)(t)$ in the above algorithm is defined using the following nested procedure (inner loop).

1. Set $k = 0$. Hold β fixed throughout the procedure. Initialize $\widehat{dH}^{(k)}(s)$ such that all jumps in the hazard have equal size.
2. Find $\widehat{dH}^{(k+1)}(s)$ according to equation (18). Increment k .
3. Repeat step (2) until

$$\ell \left(\hat{H}^{(k)}(t); \hat{\beta}^{(j)} \right) - \ell \left(\hat{H}^{(k-1)}(t); \hat{\beta}^{(j)} \right) < \varepsilon/10.$$

In practice, the value of ε had little effect on our results within a wide range. The important aspect here is that the convergence tolerance for the inner loop (used to estimate the baseline hazard) should be stricter than for the outer loop.

Consistency and asymptotic normality of these estimators is shown in Supplementary Materials Appendix E. Variance estimation is accomplished using the Hessian of the profile log-likelihood (see Appendix F for justification).

4. Simulation study

This section presents a simulation study to illustrate our method. The simulation settings are as follows. The baseline hazard was $H(t) = 2t$ (i.e., an exponential model). The true parameter vectors were $\beta_{\tau 0} = (-1, 1, -2)'$ and $\beta_{\mu 0} = (2, -1)'$. The censoring distribution was $U(0, \tau)$, $\tau = 7$. The covariates were $Z_1 \sim N(0, 1)$, $Z_2 \sim B(0.25)$, $Z_3 \sim N(0, 1)$, with $\log \eta_i = (1, z_{1i}, z_{2i})\beta_{\tau 0}$ and $\log \mu_i = (z_{3i}, z_{2i})\beta_{\mu 0}$.

We examined samples of size $n = 200, \dots, 1000$, although we give results only for $n \in \{200, 500, 1000\}$ (more detailed results are available in Supplementary Materials Appendix G). For each sample size, 1000 data sets were generated. Initial values were chosen by fitting two Cox models. We used the R function `optim()` with the BFGS method to maximize the profile likelihood function; code is available in Supplementary Materials Appendix I. Standard errors were obtained from the numerically evaluated Hessian matrix at the solution.

We simulated data under three scenarios: in the first, the model is identifiable, and the hazards for the latent and terminal events share the same shape. In this scenario, we compare our EM method with the partial likelihood estimators obtained by maximizing the partial log-likelihood in (16). This allows us to see how much information is lost by not observing the latent event.

In the second scenario, the model is still correctly specified in that the hazards for the latent and terminal events share the same shape. However, now we replace $\log \mu_i = (z_{3i}, z_{2i})\beta_{\mu 0}$ with $\log \mu_i = (z_{1i}, z_{2i})\beta_{\mu 0}$, so the model is only identifiable up to the sign of the intercept. Note that there is no issue with identifiability when the latent event is observed, as this is a consequence of the symmetric form of the marginal distribution of time to the terminal event (Section 2.3). When the latent event is observed we use the *conditional* distribution of the time to the terminal event given the latent event to obtain the partial likelihood estimators.

Finally, in the third scenario we allow the baseline hazard for the terminal event to differ from that for the latent event in order to check the robustness of the model with respect to this assumption. Specifically, if the hazard for the latent event is $\mu H(t)$, then the conditional hazard for the terminal event given the latent event is $\mathbb{1}(t > t_0) \eta [H(t)]^\alpha$ in these simulations. For this scenario, we compare the EM method with a parametric MLE for the same model assuming an exponential hazard.

4.1 Correct specification

Results for the identifiable scenario are shown in Table 1. Performance is highly dependent on sample size, with substantial bias present for small samples. However, for most parameters, this bias nearly disappears with the increase in sample size by $n = 500$. The estimated standard errors show a similar pattern: the model seems to underestimate the true variability in the estimates for the smaller sample size, but by $n = 500$ we see very good agreement between the estimated standard errors and the true standard deviation of the estimates.

Performance is adversely affected when the model is not fully identifiable, as can be seen in Table 2. Estimates obtained when the latent event is not observed display more bias and more variability than in the identifiable scenario, especially in smaller sample sizes and for the intercept parameter in particular. Furthermore, the asymptotic approximations for the standard errors are slower to converge to the true sampling variability of the estimators.

The contrast between unobserved latent event case (the EM method) and the complete-data situation (the partial likelihood estimates) is striking in both scenarios. The partial likelihood estimators are substantially less biased, and their variability is considerably lower. There is also better agreement between estimated and true standard errors in the complete-data situation for the same sample size.

The coverage probabilities seem to be relatively more robust, however. For the identifiable scenario, the coverage rates are close to the nominal level by $n = 500$, although they are slightly lower than nominal for $n = 200$. The coverage probabilities for the unidentifiable scenario deviate a bit more from the nominal level, with undercoverage for the parameters pertaining to the latent event and overcoverage for those pertaining to the terminal event. Performance is quite improved in this regard by $n = 500$, however, as was the case for the identifiable scenario. In both scenarios, the coverage probabilities for $n = 1000$ are very close to nominal, with some slight undercoverage remaining for the unidentifiable scenario.

4.2 Violation of PH assumption

All simulation settings for this scenario are as in the correctly specified scenario, with the sample size set at $n = 1000$, but now the hazard for the terminal event is $\mathbb{1}(t > t_0) \eta_i [H(t)]^\alpha$; α was varied between 0.5 and 1.5. See Figure 1 for the average of the estimated parameters across all simulated data sets for each value of α . We compared our method with a parametric version assuming an exponential hazard, which is similar in spirit to Dejardin et al. (2010), who assume a piecewise constant baseline hazard and interval-censored intermediate event.

This figure shows that the parametric model is much more affected by deviations from the PH assumption than is our semiparametric version, indicating greater robustness of the semiparametric model to violations of the proportionality assumption. While both models show some bias because of misspecification, slight deviations from the proportionality assumption do not impair our method's ability to estimate the parametric component of the model with reasonable precision.

5. Real data analysis

5.1 Latent event observed

In order to compare the proposed approach based on the EM algorithm described above with an approach considering the latent process as observed, we analyzed a data set on time to recurrence and death due to cancer in a cohort of patients with resected colon cancer Moertel et al. (1990). The goal of the study was to compare the effect of three treatments: observation, levamisole alone, and levamisole along with fluorouracil. In this case, recurrence can be considered a delayed manifestation of the latent event that marks the start of the risk of death. We identify recurrence with the latent event in our model for this analysis, although this is not entirely appropriate as it cannot be precisely observed in practice. However, our goal here is to assess robustness of the estimates with real data under a possibly misspecified model and with a possibly imperfectly defined latent event.

We analyzed the data by fitting separate Cox models to time to recurrence and time to death due to cancer, as well as models based on the joint partial likelihood and our proposed EM method. To simplify the analysis, we only included two covariates in each part of the model: dummy variables corresponding to which of the three treatment groups the subject was assigned, with the baseline category being the observation group. The results are given in Appendix H in the Supplementary Materials. The conclusions of the original study were that although levamisole showed no effect on survival or recurrence, combination therapy with fluorouracil reduced the risk of both significantly.

Our analysis confirms the effects of the treatments evaluated in the study on time to recurrence: log hazard ratios estimated in the joint Cox model are numerically very similar to those in the independent Cox fit for time to recurrence. The parameter estimates for the proposed method have the correct sign and similar magnitudes, with slightly increased standard errors, which aligns with what we would expect given intuition and our simulation results.

5.2 Latent event unobserved

To illustrate the use of the proposed method when the latent event is unobserved, we apply it to SEER registry data on prostate cancer. Specifically, we examined the survival data from the Detroit SEER registry with years of diagnosis with prostate cancer between 1983 and 2003. This was comprised of data on 47,187 men. We included two binary covariates: race (0 if white or 1 if black) and a dichotomized time of diagnosis (0 if pre- or 1 if post-1988, the year PSA screening was introduced). Of these subjects, 26.2% were black and 89.4%

were diagnosed in the PSA era (1988 or later); 7.8% died of cancer during the follow-up period.

In prostate cancer, metastasis (the latent event) must occur prior to death due to cancer. In SEER, no detailed post-treatment followup is available, so the time at which the disease becomes metastatic, even in the symptomatic sense, is unknown. The results of the conventional analysis (involving a simple Cox model fit) as well as the proposed method are shown in Table 3. We display only the positive-intercept model, but recall that due to the lack of identifiability of the sign of the intercept, we obtain the same fit to the data with a negative intercept of the same magnitude and an exchange of the roles of η and μ . This application allows us to illustrate the resolution of the problem of identifiability in a practical situation. We emphasize that this choice of model is possible only through the use of external information; no statistical justification can be made, as both models fit the data equally well.

The rationale for our choice of the positive intercept model is that it results in the correct sign for the effect of PSA screening on time to death. In the positive intercept model, this coefficient is negative, which agrees with the Cox model's estimate in sign if not magnitude. Moreover, it is an accepted scientific view that PSA screening prolongs time to death, if only due to the artifact of lead-time bias.

Interpretation of the coefficient estimates is similar in our model to what it is in the Cox model. However, instead of marginal log hazard ratios of time to death, the terminal event part of our model produces estimates of the conditional log hazard ratios of time to death, given time to onset of metastasis. For example, while black patients have a risk of death roughly 40% higher than white patients according to the Cox model, given time to onset of metastasis, this risk is only 35% higher. The difference in effect sizes is more pronounced for the effect of PSA screening: marginally, it reduces risk of death by 57%, but given time to onset of metastasis, the reduction is 84%.

Equations (5) and (6), the conditional survival functions for onset of metastasis given time to death or censoring, allow us to produce the plots shown in Figure 2. An immediately evident feature of these curves is their ordering, with the PSA-screened population below the non-screened population, indicating an earlier onset of metastasis for these subjects. This may be explained with reference to the selection effect caused by screening. Under PSA screening, tumors are detected earlier and treated, in some sense removing these cases from the population. Due to the length bias (Zelen and Feinleib, 1969), these tumors will be generally less aggressive, so the remainder—i.e., the cases which would be depicted in these plots, conditioning on death—will be relatively more aggressive, with earlier metastasis.

6. Discussion

This paper presents a method for jointly modeling time to a latent event and time to the terminal event, in a semiparametric framework, when the latent event is unobservable. Our approach involves an EM algorithm for estimating the baseline hazard, the derivation of which constitutes the chief methodological contribution of our work. EM represents a stable,

computationally efficient solution that handles the curse of dimensionality in a closed form and maximizes the likelihood even if the model is non-identifiable. The method is generalizable, in the sense that it is not specific to a certain data structure, but can instead be used with any survival data for which there is interest in factors affecting time to an unobserved event which is known to precipitate an observed event.

An alternative to our method is a weighted Breslow-type estimator (Chen, 2009) that is also asymptotically fully efficient, although this estimator needs an extension to our class of problems. However, the weighted Breslow method generally does not enjoy the property of monotonic convergence characteristic of the EM approach. In fact, EM will converge even in the unidentifiable case with unrestricted sign of the intercept term. This property makes the EM approach especially suitable for our problem.

The issue of identifiability when our model contains the same set of covariates in both the latent and terminal event hazards adversely affects numerical performance and presents a case of nonstandard interpretation. In essence, in this situation our method produces not one but two possible models, each with its own interpretation. Absent external considerations that exclude one of the two models, both models need to be reported as a simple alternative that represents a correct expression of the inference uncertainty inherited from the structure of the data. As an example of such external considerations, an examination of the sign of the estimated coefficients and the expected direction of effect for each part of the model provided guidance as to the correct choice of model in our analysis of the prostate cancer data. Often, in practice, the sets of significant covariates in the two predictors of the final model are different, and interpretation becomes unambiguous.

In order to assess the model fit, Dejardin et al. (2010) in their Discussion recommend comparison of Kaplan-Meier plots for the marginal survival functions of the observed terminal event with the marginal survival functions obtained using their proposed multi-state model. We used a similar approach and found that the Kaplan-Meier plots alongside plots of the model-predicted marginal survival functions are very similar (not shown). This provides a simple graphical check of the adequacy of the model fit.

We have shown in detail why we are able to use the Hessian of the profile log-likelihood as an estimate of the information matrix (see Supplementary Materials Appendix F). Convergence of the estimated covariance matrix to the truth in our simulation studies is nonetheless rather slow. This is typical of models that invoke a latent structure. We also have used a numerical approximation to the Hessian of the profile log-likelihood. The methods of Tsodikov and Garibotti (2007) can be used to obtain the exact profile information matrix if it is suspected that this approximation may be inadequate.

The assumption of proportional hazards between the latent and terminal events, also made by Dejardin et al. (2010), is difficult to verify in general without some parametric assumptions. We see the parametric structural assumptions as a necessary tradeoff: when less data is available, stronger assumptions must be made to achieve useful statistical inference. In a simulation study, we allowed the hazard for the terminal event to differ from that for the latent event and observed that the model is robust to minor deviations from this

assumption, and greatly outperforms its parametric counterpart in this respect. Nevertheless, relaxation of the proportional hazards assumption provides an important direction for future research. Linking the two hazards by a parametric transformation model is one such opportunity that can be handled within the general framework of this paper.

Mapping unobserved mathematical constructs to biological entities is necessarily fuzzy. While in the prostate cancer example patients die of metastases rather than of the local tumor growth, the onset of metastasis, representing the latent event that marks the starting point for the risk of death, is still a vague definition from a biological standpoint. However, until more precise measurements of the precursor events that make cancer lethal become available, latent event models will be a useful tool that capture the general structural understanding of the disease progression.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research was supported by the grant U01CA199338 (CISNET) from the National Cancer Institute. Additional support was provided by “Modeling to Improve Prostate Cancer Outcomes Across Diverse Populations” grant 1U01CA199338 (CISNET). The authors are also grateful to two anonymous referees for their helpful comments and suggestions.

References

- Chen YH. Weighted Breslow-type and maximum likelihood estimation in semi-parametric transformation models. *Biometrika*. 2009; 96:591–600.
- Dejardin D, Lesaffre E, Verbeke G. Joint modeling of progression-free survival and death in advanced cancer clinical trials. *Statistics in Medicine*. 2010; 29:1724–1734. [PubMed: 20572123]
- Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B (Methodological)*. 1977; 39:1–38.
- Frydman H. Semiparametric estimation in a three-state duration-dependent Markov model from interval-censored observations with application to AIDS data. *Biometrics*. 1995; 51:502–511. [PubMed: 7662839]
- Frydman H, Szarek M. Nonparametric estimation in a Markov “illness-death” process from interval censored observations with missing intermediate transition status. *Biometrics*. 2009; 65:143–151. [PubMed: 18505421]
- Gjessing HK, Aalen OO, Hjort NL. Frailty models based on Lévy processes. *Advances in Applied Probability*. 2003; 35:532–550.
- Hu C, Tsodikov A. Joint modeling approach for semicompeting risks data with missing nonterminal event status. *Lifetime Data Analysis*. 2014a; 20:563–583. [PubMed: 24430204]
- Hu C, Tsodikov A. Semiparametric regression analysis for time-to-event marked endpoints in cancer studies. *Biostatistics*. 2014b; 15:513–525. [PubMed: 24379192]
- Kalbfleisch, JD., Prentice, RL. *The Statistical Analysis of Failure Time Data*. 2. Wiley; 2002.
- Lin DY, Sun W, Ying Z. Nonparametric estimation of the gap time distributions for serial events with censored data. *Biometrika*. 1999; 86:59–70.
- Moertel CG, Fleming TR, Macdonald JS, Haller DG, Laurie JA, Goodman PJ, Ungerleider JS, Emerson WA, Tormey DC, Glick JH, Veeder MH, Mailliard JA. Levamisole and fluorouracil for adjuvant therapy of resected colon carcinoma. *New England Journal of Medicine*. 1990; 322:352–358. [PubMed: 2300087]

- Murphy SA, van der Vaart AW. On profile likelihood. *Journal of the American Statistical Association*. 2000; 95:449–465.
- Sternberg MR, Satten GA. Discrete-time nonparametric estimation for semi-Markov models of chain-of-events data subject to interval censoring and truncation. *Biometrics*. 1999; 55:514–522. [PubMed: 11318208]
- Tsodikov A. Semiparametric models: a generalized self-consistency approach. *Journal of the Royal Statistical Society, Series B (Methodological)*. 2003; 65:759–774.
- Tsodikov A, Garibotti G. Profile information matrix for nonlinear transformation models. *Lifetime Data Analysis*. 2007; 13:139–159. [PubMed: 17021958]
- Tsodikov AD, Asselain B, Fourque A, Hoang T, Yakovlev AY. Discrete strategies of cancer post-treatment surveillance: Estimation and optimization problems. *Biometrics*. 1995; 51:437–447. [PubMed: 7662836]
- Vaupel JW, Manton KG, Stallard E. The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography*. 1979; 16:439–454. [PubMed: 510638]
- Zelen M, Feinleib M. On the theory of screening for chronic diseases. *Biometrika*. 1969; 56:601–614.

Figure 1.

Simulation results for model misspecification: each panel shows a different parameter in the model, with the top row giving parameters associated with the observed terminal event and the first two panels in the bottom row showing the parameters associated with the unobserved latent event.

Figure 2.

Conditional survival functions for onset of metastasis given observed data (time to death and censoring indicator)—positive-intercept model. Top row is for a hypothetical subject censored at the times indicated, while the bottom row is for a subject who dies at these times.

Table 1

Simulation results, identifiable scenario. This table shows the bias, average standard error (ASE), empirical standard deviation (ESD), and 95% confidence interval coverage probability (CP) across 1000 simulated data sets at each sample size. The two methods compared maximize the partial log-likelihood (14) when the latent event is observed and the profile log-likelihood (8) when it is unobserved. All values (excluding CP) appearing in the table are normalized by dividing by the absolute true parameter value to facilitate comparisons of estimates between coefficients.

Sample size	Covariate	Latent observed				Latent unobserved			
		Bias	ASE	ESD	CP	Bias	ASE	ESD	CP
200	(Intercept)	-0.016	0.170	0.171	0.961	-0.037	0.496	0.606	0.930
	$Z_1 \sim N(0, 1)$	0.010	0.119	0.123	0.946	0.049	0.163	0.182	0.940
	$Z_2 \sim B(0.25)$	-0.013	0.179	0.186	0.952	0.020	0.266	0.341	0.919
	$Z_3 \sim N(0, 1)$	0.007	0.068	0.069	0.950	0.057	0.223	0.278	0.923
	$Z_2 \sim B(0.25)$	-0.006	0.204	0.200	0.951	-0.027	0.975	1.231	0.914
500	(Intercept)	-0.005	0.106	0.107	0.953	-0.010	0.300	0.312	0.941
	$Z_1 \sim N(0, 1)$	0.005	0.074	0.075	0.941	0.016	0.098	0.099	0.956
	$Z_2 \sim B(0.25)$	-0.003	0.110	0.108	0.958	0.003	0.156	0.164	0.943
	$Z_3 \sim N(0, 1)$	0.006	0.043	0.044	0.933	0.019	0.134	0.141	0.935
	$Z_2 \sim B(0.25)$	0.002	0.127	0.131	0.944	0.006	0.589	0.650	0.955
1000	(Intercept)	-0.003	0.075	0.076	0.950	0.005	0.207	0.213	0.942
	$Z_1 \sim N(0, 1)$	0.003	0.051	0.052	0.950	0.010	0.068	0.070	0.952
	$Z_2 \sim B(0.25)$	-0.000	0.077	0.080	0.932	-0.001	0.109	0.113	0.945
	$Z_3 \sim N(0, 1)$	0.001	0.030	0.032	0.939	0.005	0.092	0.096	0.943
	$Z_2 \sim B(0.25)$	0.000	0.090	0.089	0.951	0.005	0.408	0.412	0.952

Table 2

Simulation results, unidentifiable scenario. This table shows the bias, average standard error (ASE), empirical standard deviation (ESD), and 95% confidence interval coverage probability (CP) across 1000 simulated data sets at each sample size. The two methods compared maximize the partial log-likelihood (14) when the latent event is observed and the profile log-likelihood (8) when it is unobserved. All values (excluding CP) appearing in the table are normalized by dividing by the absolute true parameter value to facilitate comparisons of estimates between coefficients.

Sample size	Covariate	Latent observed				Latent unobserved			
		Bias	ASE	ESD	CP	Bias	ASE	ESD	CP
200	(Intercept)	-0.023	0.176	0.177	0.940	0.301	1.238	0.813	0.991
	$Z_1 \sim N(0, 1)$	0.019	0.142	0.141	0.948	-0.038	0.359	0.250	0.969
	$Z_2 \sim B(0.25)$	-0.014	0.163	0.167	0.952	0.121	0.335	0.454	0.930
	$Z_1 \sim N(0, 1)$	0.012	0.069	0.069	0.949	0.018	0.284	0.353	0.899
	$Z_2 \sim B(0.25)$	-0.010	0.204	0.204	0.951	-0.209	1.275	1.528	0.887
500	(Intercept)	-0.004	0.110	0.110	0.944	0.115	0.725	0.757	0.973
	$Z_1 \sim N(0, 1)$	0.009	0.088	0.089	0.951	-0.017	0.200	0.175	0.966
	$Z_2 \sim B(0.25)$	-0.011	0.100	0.105	0.941	0.017	0.161	0.250	0.935
	$Z_1 \sim N(0, 1)$	0.005	0.042	0.043	0.947	0.020	0.175	0.263	0.931
	$Z_2 \sim B(0.25)$	-0.001	0.127	0.128	0.954	0.031	0.815	1.039	0.933
1000	(Intercept)	-0.002	0.077	0.080	0.942	0.095	0.534	0.545	0.928
	$Z_1 \sim N(0, 1)$	0.005	0.062	0.063	0.951	-0.019	0.147	0.145	0.930
	$Z_2 \sim B(0.25)$	0.000	0.070	0.071	0.950	0.004	0.102	0.113	0.944
	$Z_1 \sim N(0, 1)$	0.003	0.030	0.032	0.937	0.009	0.122	0.132	0.930
	$Z_2 \sim B(0.25)$	-0.002	0.090	0.092	0.952	0.029	0.547	0.645	0.944

Table 3

Parameter estimates (standard errors) from analysis of SEER prostate cancer data. The Cox model estimates are shown for purposes of comparison with the estimates for the η part of the joint model, which pertains to time to death due to cancer given metastasis.

	Parameter	Cox	Joint
<i>Death</i>	(Intercept)	—	2.136 (0.233)
	Black	0.338 (0.036)	0.301 (0.114)
	Dx post-1988	−0.845 (0.038)	−1.813 (0.381)
<i>Onset</i>	Black	—	0.137 (0.097)
	Dx post-1988	—	0.401 (0.357)