Ann Inst Stat Math (2011) 63:117–134
DOI 10.1007/s10463-008-0212-2

# Additive risk model for current status data with a cured subgroup

**Shuangge Ma**

Received: 8 May 2008 / Revised: 17 September 2008 / Published online: 20 December 2008


**Abstract** Current status data arise when the exact timing of an event cannot be observed, and the only available information is whether or not the event has occurred at a random censoring time point. We consider current status data with a cured subgroup, where subjects in this subgroup are not susceptible to the event of interest. We model the cure probability using a generalized linear model with a known link function. For subjects susceptible to the event, we model their survival hazard using a partly linear additive risk model. We show that the penalized maximum likelihood estimate of the parametric regression coefficient is $\sqrt{n}$ consistent, asymptotically normal and efficient. The nonparametric cumulative baseline function and nonparametric covariate effect can be estimated with the $n^{1/3}$ convergence rate. We propose inference using the weighted bootstrap. Simulations study is employed to assess finite sample performance of the proposed estimate. We analyze the Calcification study using the proposed approach.

**Keywords** Additive risk model · Cure model · Current status data · M-estimator

## 1 Introduction

Current status data, also known as case I interval censored data, arise naturally in many fields of studies. With current status data, the event time of interest cannot be accurately observed—the only available information is whether the event has occurred, i.e, the event status, at a random censoring time point. Examples of current status data include the demographic study of Diamond and McDonald (1991), AIDS study of Shiboski and Jewell (1992), tumor clinical study of Gart et al. (1986), and many

S. Ma (✉)
Department of Epidemiology and Public Health, Yale University, New Haven, CT 06510, USA
e-mail: shuangge.ma@yale.edu

others. When there is no covariate, the survival of current status data can be modeled with nonparametric functions (Groeneboom and Wellner 1992; van der Vaart 1998). When there exist covariates, models investigated include the linear Cox model (Huang 1996), additive risk model (Lin 1998; Martinussen and Scheike 2002), partly linear accelerated failure time (AFT) model (Xue et al. 2004), and partly linear transformation model (Ma and Kosorok 2005a). In the aforementioned studies, it has been assumed that, if the followup is long enough, all the subjects will eventually experience the event of interest, i.e, if we denote $S$ as the survival function, then $S(+\infty) = 0$.

Recent studies, such as Yu et al. (2001) and Thompson and Chhikara (2003), show that, for interval censored data, it may be necessary to consider generalizations of the commonly used survival models and allow for a cured subgroup. Here, individuals in the cured subgroup are not susceptible to the event. Such models have been generally referred to as "cure models". Under the cure model, $S(+\infty) > 0$. With right censored survival data, studies of cure models include Farewell (1986), Lu and Ying (2004), Peng and Dear (2000), Li et al. (2001), and Chen et al. (2004), and many others. With interval censored data and a cured subgroup, Lam and Xue (2005) assumes the partly linear AFT model for the event time and considers a sieve maximum likelihood approach; Ma (2007) assumes the partly linear Cox model and considers a penalized approach; Thompson and Chhikara (2003) proposes a parametric model for the event time and a Bayesian estimator. When there exist correlations among subjects, a parametric model for the event time is investigated in Bellamy et al. (2004); When there exist spatial associations among subjects, a frailty model with a Bayesian approach is proposed in Banerjee and Carlin (2004).

In this article, we investigate current status data with a cured subgroup. We assume that the cure probability satisfies a generalized linear model with a known link function. For susceptible subjects, the event time is modeled using a partly linear additive risk model. As a flexible alternative to the Cox model, the additive risk is adopted when the covariates contribute to the hazard function in an additive manner. Previous studies have shown that the additive risk model has sound biological basis and can be very useful when the Cox model fails (Lin 1998; Martinussen and Scheike 2002). The assumed partly linear additive risk model is more flexible than the linear additive risk model in Lin (1998) and shares similar spirits as the partly linear Cox model in Huang (1999) and the partly linear transformation model in Ma and Kosorok (2005a). Although the aforementioned studies have demonstrated great values of the additive risk model, such a model has not been used to describe data with a cured subgroup, not even for right censored data. Our study targets to fill in this missing piece. The modeling strategy and technical tools used in this study have certain overlaps with Lam and Xue (2005), Ma (2007), Huang (1996), and Ma and Kosorok (2005a). However, the unique characteristics of the additive risk model make this study fundamentally different from those. The new additive risk model, together with its sound theoretical properties, will provide more insights into analysis of heterogeneous interval censored data.

The article is organized as follows. In Sect. 2, we first introduce the data structure. The partly linear additive risk model is introduced in Sect. 3. Asymptotic properties of the penalized maximum likelihood estimate (PMLE) are then established. Numerical studies, including simulation studies and analysis of the Calcification data, are

provided in Sect. 4. The article concludes with discussions in Sect. 5. Proofs are provided in the Appendix.

## 2 Data and model settings

Denote $T$ as the event time of interest and $C$ as the random censoring time. For simplicity of notations, we assume only two covariates $Z_1$ and $Z_2$. Denote $Z = (Z_1, Z_2)'$. For current status data, one observation consists of $X = (C, \delta = I(T \leq C), Z_1, Z_2)$. To account for the possibility of cure, we introduce the unobservable cure indicator $U$: $U = 0$ if the subject is cured (i.e., $T = \infty$) and $U = 1$ otherwise. We model the cure probability using a generalized linear model with a known link function. We are particularly interested in the logistic link function (Lu and Ying 2004; Lam and Xue 2005), where

$$P(U = 1|Z) = \phi(\alpha'\tilde{Z}) = \frac{\exp(\alpha'\tilde{Z})}{1 + \exp(\alpha'\tilde{Z})}. \tag{1}$$

Here $\alpha$ is the unknown regression coefficient, $\alpha'$ is the transpose of $\alpha$, and $\tilde{Z} = (1, Z')'$.

For subjects with $U = 1$, we model the survival hazard of event time $T$ using the partly linear additive risk model:

$$\lambda(T|Z) = \lambda(T) + f(Z), \quad \text{where } f(Z) = \beta Z_1 + h(Z_2). \tag{2}$$

Here $\lambda(T|Z)$ is the conditional hazard function, $\lambda(T)$ is the unknown baseline hazard function, $\beta$ is the parametric regression coefficient, and $h$ is the smooth, nonparametric covariate effect. The cumulative hazard function is:

$$\Lambda(T|Z) = \Lambda(T) + f(Z)T.$$

Great flexibility is introduced in (2) by allowing for nonparametric covariate effect $h$. A special example of model (2) is the linear additive risk model with $h(Z_2) = \beta_2 Z_2$ (Lin 1998). We note that, a more general choice of $f$ is $f(Z) = f(Z_1, Z_2)$, where $f$ is the unknown, nonparametric function. With such a general formulation, smoothing techniques such as spline or kernel estimations will be needed. In this article, we make the assumption of a semiparametric model. Specifically, we assume that there is no "interaction" term of $Z_1$ and $Z_2$. Such an assumption can be restrictive compared to the general formulation. However, it is still flexible enough, and can accommodate many practical scenarios. In addition, the semiparametric model being considered may have more intuitive interpretations than the nonparametric model, and can hence be preferred in many practical studies.

If we assume the event time and censoring are conditionally independent, then for a single observation $X$, the log-likelihood (up to a constant) is equal to

$$\begin{aligned} l(X; \alpha, f, \Lambda) = {} & \delta \log(\phi(\alpha'\tilde{Z})) + \delta \log[1 - \exp(-\Lambda(C) - f(Z)C)] \\ & + (1 - \delta) \log\{1 - \phi(\alpha'\tilde{Z})[1 - \exp(-\Lambda(C) - f(Z)C)]\}. \end{aligned} \tag{3}$$

## 3 Penalized maximum likelihood estimation

In the estimation, constraints on the nonparametric covariate effect $h$ are needed. Otherwise, over-fitting may occur. In our study, we assume that $h$ is smooth, more specifically, a spline function. See assumption A4 below. We propose using penalty to control the smoothness of $h$, whereas in Lam and Xue (2005), the sieve approach is used. An advantage of the penalized approach is that the degree of smoothness can be easily controlled by a single data-dependent parameter—the penalty term.

Assume that $n$ iid observations $X_1 = (C_1, \delta_1, Z_{11}, Z_{21}), \ldots, X_n = (C_n, \delta_n, Z_{1n}, Z_{2n})$ are available. We consider the penalized maximum likelihood estimate (PMLE)

$$(\hat{\alpha}, \hat{\beta}, \hat{h}, \hat{\Lambda}) = \arg \max_{\alpha,\beta,h,\Lambda} \mathrm{P}_n l(X) - \lambda_n^2 J^2(h), \tag{4}$$

where $\lambda_n$ is the data-dependent tuning parameter, $J^2(h)$ is the penalty on smoothness defined as $J^2(h) = \int_{Z_2} (h^{(s_0)}(Z_2))^2 dZ_2$, $h^{(s_0)}$ is the $s_0$th derivative of $h$, and $\mathrm{P}_n$ is the empirical measure. In practical data analysis, it is commonly assumed that $s_0 = 2$ (Wahba 1990).

Denote the unknown true value of $(\alpha, \beta, h, \Lambda)$ as $(\alpha_0, \beta_0, h_0, \Lambda_0)$. We first make the following assumptions:

A1. (1) $T$ and $C$ are conditionally independent given $Z$. (2) The support of $C$ is an interval $[l_C, u_C]$ with $0 \le l_C < u_C < \infty$.

A2. (1) $Z$ belongs to a bounded subset of $\mathbb{R}^2$. (2) The parametric parameter $(\alpha_0, \beta_0)$ belongs to a compact subset of $\mathbb{R}^4$. (3) For any $\alpha \neq \alpha_0$, $Pr(\alpha' \tilde{Z} \neq \alpha_0' \tilde{Z}) > 0$. For any $\beta \neq \beta_0$, $Pr(\beta Z_1 \neq \beta_0 Z_1) > 0$.

A3. For $l_C \le T \le u_C$: (1) there exists a constant $M$, such that $0 < \frac{1}{M} < \Lambda_0(T) < M < \infty$. (2) $\Lambda_0$ has strictly positive first order derivative.

A4. (1) $h_0 \in \mathbb{H}_{s_0}$, the Sobolev space indexed by the order of derivative $s_0 \ge 1$; (2) $\mathrm{P}(h_0(Z_2)) = 0$, where P is the expectation.

Define the distance between $(\alpha, \beta, h, \Lambda)$ and $(\alpha_0, \beta_0, h_0, \Lambda_0)$ as

$$d((\alpha, \beta, h, \Lambda), (\alpha_0, \beta_0, h_0, \Lambda_0)) = ||\alpha - \alpha_0|| + ||\beta - \beta_0|| + ||h - h_0||_2 + ||\Lambda - \Lambda_0||_2,$$

where $||h - h_0||_2^2 = \int_{Z_2} (h(Z_2) - h_0(Z_2))^2 dZ_2$ and $||\Lambda(c) - \Lambda_0(c)||_2^2 = \int_{l_C}^{u_C} (\Lambda(c) - \Lambda_0(c))^2 dc$. Beyond assumptions A1–A4, we also assume

A5. For any $(\alpha, \beta, h, \Lambda)$ satisfying assumptions A1–A4,

$$\mathrm{P}[l(\alpha, \beta, h, \Lambda) - l(\alpha_0, \beta_0, h_0, \Lambda_0)] \le -K_1 d^2((\alpha, \beta, h, \Lambda), (\alpha_0, \beta_0, h_0, \Lambda_0)),$$

with a fixed constant $K_1 > 0$.

Similar assumptions have been made in Huang (1996) and van der Vaart (1998). We note that, with the compactness assumptions, the actual bounds may remain

unknown—they are not needed in the theoretical proof or computation. In addition, we assume the tuning parameter

A6. $\lambda_n = O_p(n^{-1/3})$.

*Remark 1* For subjects with $U = 1$, their survival function should satisfy $S(\infty) = 0$. We note that with the compactness assumptions A1–A4, $S(u_C) < \infty$. This does not contradict the $S(\infty) = 0$ assumption, as $S$ cannot be estimated for $T > u_C$. For practical data analysis, we can set $\hat{\Lambda} = \infty$ for $T > \max\{C_n\}$. Related practical issues have been discussed in Li et al. (2001). If we can conclude $h = h_0$ when $\{h \neq h_0\}$ is a measure zero set, then for $l_C \leq T \leq u_C$, the proposed model is identifiable. Proof follows from Li et al. (2001). We note that by assuming specific forms of the link function $\phi$ and the survival function, the identifiability assumptions are weaker than those in Li et al. (2001).

*Remark 2* Compactness assumptions are made in A1–A4. Under those assumptions and A6, the PMLE defined in (4) exists. For any finite $n$, we do not assume the uniqueness of the PMLE. However under assumptions A5 and A6, the PMLE is asymptotically unique when $n \to \infty$.

### 3.1 Finite sample properties

Let $C_{(1)}, \ldots, C_{(n)}$ be the ordered $C_1, \ldots, C_n$. Let $\delta_{(i)}, Z_{(1i)}, Z_{(2i)}$ be the indicator and covariates corresponding to $C_{(i)}$. Since only the values of $\Lambda$ at $C_{(i)}$ matter in the log-likelihood function, we set the PMLE $\hat{\Lambda}$ as the right-continuous non-decreasing step function with jumps only at $C_{(i)}$. Following Groeneboom and Wellner (1992), we assume $\delta_{(1)} = 1$ and $\delta_{(n)} = 0$, since the terms associated with $\delta_{(1)} = 0$ and $\delta_{(n)} = 1$ have zero contributions to the likelihood.

**Lemma 1** *The PMLE defined in* (4) *satisfies*

$$\left.\frac{\partial P_n l}{\partial \alpha}\right|_{\alpha=\hat{\alpha},\beta=\hat{\beta},h=\hat{h},\Lambda=\hat{\Lambda}} = 0, \quad \left.\frac{\partial P_n l}{\partial \beta}\right|_{\alpha=\hat{\alpha},\beta=\hat{\beta},h=\hat{h},\Lambda=\hat{\Lambda}} = 0, \tag{5}$$

$$\begin{aligned}
&\sum_{j\geq i}\Bigg(\frac{\delta_{(j)}}{1-\exp(-\hat{\Lambda}_{(j)}-(\hat{\beta}Z_{(1j)}+\hat{h}(Z_{2j}))C_{(j)})} \\
&\quad - \frac{(1-\delta_{(j)})\phi}{1-\phi[1-\exp(-\hat{\Lambda}_{(j)}-(\hat{\beta}Z_{(1j)}+\hat{h}(Z_{2j}))C_{(j)})]}\Bigg) \\
&\quad \times \exp(-\hat{\Lambda}_{(j)}-(\hat{\beta}Z_{(1j)}+\hat{h}(Z_{2j}))C_{(j)}) \leq 0,
\end{aligned} \tag{6}$$

$$\begin{aligned}
&\sum_{i=1}^{n}\Bigg(\frac{\delta_{(i)}}{1-\exp(-\hat{\Lambda}_{(i)}-(\hat{\beta}Z_{(1i)}+\hat{h}(Z_{2i}))C_{(i)})} \\
&\quad - \frac{(1-\delta_{(i)})\phi}{1-\phi[1-\exp(-\hat{\Lambda}_{(i)}-(\hat{\beta}Z_{(1i)}+\hat{h}(Z_{2i}))C_{(i)})]}\Bigg) \\
&\quad \times \exp(-\hat{\Lambda}_{(i)}-(\hat{\beta}Z_{(1i)}+\hat{h}(Z_{2i}))C_{(i)})\hat{\Lambda}_{(i)} = 0,
\end{aligned} \tag{7}$$

*for* $i = 1, \ldots, n$. *Equations* (5) *hold following the definition of the PMLE. Equations* (6) *and* (7) *can be proved by following Proposition* 1.1 *of* Groeneboom and Wellner (1992). *The proof is omitted here.*

*Remark 3* It can be shown that $\hat{h}$ defined in (4) is a spline function. Specifically, suppose that $\tilde{h}$ maximizes the penalized log-likelihood function. Then there exists a spline function $\hat{h}$, such that $\hat{h}(Z_{2i}) = \tilde{h}(Z_{2i})$ for $i = 1, \ldots, n$ and $J(\hat{h}) \leq J(\tilde{h})$ (Wahba 1990).

### 3.2 Consistency and rate of convergence

**Lemma 2** *Under assumptions A1–A6,*

$$d((\hat{\alpha}, \hat{\beta}, \hat{h}, \hat{\Lambda}), (\alpha_0, \beta_0, h_0, \Lambda_0)) = O_p(n^{-1/3}) \quad \textit{and} \quad J(\hat{h}) = O_p(1).$$

With current status data, Groeneboom and Wellner (1992) and Huang (1996) have proved that, under assumptions similar to A1–A5, the best possible convergence rate for estimates of the nonparametric baseline function is $n^{1/3}$. This is considerably slower than the $n^{1/2}$ rate with right censored data, due to excessive censoring. Lemma 2 shows that the optimal convergence rate for estimating $\Lambda$ can in fact be achieved. In Lam and Xue (2005), it is shown that the nonparametric covariate effect can be estimated at the optimal $n^{s_0/(2s_0+1)}$ convergence rate. However, under the partly linear additive risk model, $h_0$ can only be estimated at the much slower $n^{1/3}$ rate. The overall entropy is driven by the entropy of $\Lambda$, resulting in an overall convergence rate of $n^{1/3}$. We note that assumption A6 is different from the commonly assumed $\lambda_n = O_p(n^{-s_0/2s_0+1})$ (Wahba 1990). Changing this assumption will not improve the convergence rate of $\hat{h}$, as can be seen from the proof.

### 3.3 Fisher information

For semiparametric models, $\sqrt{n}$ consistency and asymptotic normality of the maximum likelihood estimates usually require existence and non-singularity of the information matrix. Due to the presence of the second nonparametric parameter $h$, standard information calculation based on the orthogonal projection cannot be used. As an alternative, we apply the non-orthogonal projection (Sasieni 1992).

The score functions for $\alpha$ and $\beta$ are simply the derivatives of the log-likelihood function:

$$\dot{l}_\alpha = \left[\frac{\delta}{\phi} - \frac{(1-\delta)g}{1-\phi g}\right]\dot{\phi}\tilde{Z}, \quad \dot{l}_\beta = \left[\frac{\delta}{g} - \frac{(1-\delta)\phi}{1-\phi g}\right]\exp(-\Lambda - (\beta Z_1 + h(Z_2))C)Z_1 C,$$

where $g(\alpha, \beta, h, \Lambda) = 1 - \exp[-\Lambda - (\beta Z_1 + h(Z_2))C]$. Denote $\dot{l}_{\alpha\beta} = (\dot{l}_\alpha, \dot{l}_\beta)'$.

For $\eta \sim 0$, consider a small perturbation of $h$, $h_\eta = h + \eta\xi(Z_2)$, such that $h_\eta$ still satisfies assumption A4. Denote the space generated by such $\xi$ as $\mathbb{B}$. We have $\frac{\partial h_\eta}{\partial \eta}|_{\eta=0} = \xi(Z_2)$. Thus the score operator for $h$ is

$$\dot{l}_h(\xi) = \left[\frac{\delta}{g} - \frac{(1-\delta)\phi}{1-\phi g}\right]\exp(-\Lambda - (\beta Z_1 + h(Z_2))C)\xi.$$

Denote $\mathbb{A} = \{a : \int_{l_C}^{u_C} a(C)dC = 0;\ a \in L_2(P)$; for $u$ small enough $\Lambda_u = \Lambda + ua$ satisfies assumption $A3\}$. The score operator for $\Lambda$ is:

$$\dot{l}_\Lambda(a) = -\left[\frac{\delta}{g} - \frac{(1-\delta)\phi}{1-\phi g}\right]\exp(-\Lambda - (\beta Z_1 + h(Z_2))C)a = \tilde{l}_\Lambda a.$$

Consider the following three-step projection.

*Step 1*. We first project $\dot{l}_{\alpha\beta}$ onto the space generated by $\dot{l}_\Lambda$. We need to find the "direction" $a^* \in \mathbb{A}$ such that $\dot{l}_{\alpha\beta} - \dot{l}_\Lambda(a^*) \perp \dot{l}_\Lambda(a)$ for all $a \in \mathbb{A}$. This is equivalent to require $\mathrm{P}[(\dot{l}_{\alpha\beta} - \tilde{l}_\Lambda a^*)\tilde{l}_\Lambda a] = 0$. Applying the standard projection, we can see that if $a^* = \frac{\mathrm{P}(\dot{l}_{\alpha\beta}\tilde{l}_\Lambda|C)}{\mathrm{P}(\tilde{l}_\Lambda\tilde{l}_\Lambda|C)}$, then this requirement is satisfied. Hence we have

$$\dot{l}_{\alpha\beta} - \dot{l}_\Lambda(a^*) = \dot{l}_{\alpha\beta} - \tilde{l}_\Lambda\frac{\mathrm{P}(\dot{l}_{\alpha\beta}\tilde{l}_\Lambda|C)}{\mathrm{P}(\tilde{l}_\Lambda\tilde{l}_\Lambda|C)}. \tag{8}$$

*Step 2*. We now project $\dot{l}_h(\xi)$ onto the space generated by $\dot{l}_\Lambda$, using calculations similar to those in Step 1. Denote the least favorable direction as $b^* \in \mathbb{A}$ and we have

$$\dot{l}_h(\xi) - \dot{l}_\Lambda(b^*) = \dot{l}_h(\xi) - \tilde{l}_\Lambda\frac{\mathrm{P}(\dot{l}_h\tilde{l}_\Lambda|C)}{\mathrm{P}(\tilde{l}_\Lambda\tilde{l}_\Lambda|C)}. \tag{9}$$

*Step 3*. We project the space generated by $\dot{l}_{\alpha\beta} - \dot{l}_\Lambda(a^*)$ onto the space generated by $\dot{l}_h(\xi) - \dot{l}_\Lambda(b^*)$, which is equivalent to find $\xi^* \in \mathbb{B}$ such that

$$\mathrm{P}\left\{\left(\left[\dot{l}_{\alpha\beta} - \tilde{l}_\Lambda\frac{\mathrm{P}(\dot{l}_{\alpha\beta}\tilde{l}_\Lambda|C)}{\mathrm{P}(\tilde{l}_\Lambda\tilde{l}_\Lambda|C)}\right] - \left[\dot{l}_h(\xi^*) - \tilde{l}_\Lambda\frac{\mathrm{P}(\dot{l}_h(\xi^*)\tilde{l}_\Lambda|C)}{\mathrm{P}(\tilde{l}_\Lambda\tilde{l}_\Lambda|C)}\right]\right) \times \left(\dot{l}_h(\xi) - \tilde{l}_\Lambda\frac{\mathrm{P}(\dot{l}_h(\xi)\tilde{l}_\Lambda|C)}{\mathrm{P}(\tilde{l}_\Lambda\tilde{l}_\Lambda|C)}\right)\right\} = 0, \tag{10}$$

for any $\xi \in \mathbb{B}$. Denote $\dot{l} = \left[\dot{l}_{\alpha\beta} - \tilde{l}_\Lambda\frac{\mathrm{P}(\dot{l}_{\alpha\beta}\tilde{l}_\Lambda|C)}{\mathrm{P}(\tilde{l}_\Lambda\tilde{l}_\Lambda|C)}\right] - \left[\dot{l}_h(\xi^*) - \tilde{l}_\Lambda\frac{\mathrm{P}(\dot{l}_h(\xi^*)\tilde{l}_\Lambda|C)}{\mathrm{P}(\tilde{l}_\Lambda\tilde{l}_\Lambda|C)}\right]$. We assume

A7. There exists $\xi^* \in \mathbb{B}$ such that (10) is satisfied and $I = \mathrm{P}(\dot{l}\dot{l}')$ is positive definite and component-wise bounded.

**Lemma 3** *Under assumptions* A1–A5 *and* A7, *$I$ is the efficient information matrix for* $(\alpha, \beta)$.

### 3.4 Asymptotic normality

**Lemma 4** *Under assumptions* A1–A7, $\sqrt{n}(\hat{\alpha} - \alpha_0, \hat{\beta} - \beta_0) \to_d N(0, I^{-1})$.

Lemma 4 suggests that the parametric regression parameters are still $\sqrt{n}$ estimable, despite the slow convergence rate of the baseline and nonparametric covariate effect estimates. Moreover, the proposed PMLE is asymptotically efficient, in the sense that any regular estimator would have asymptotic covariance matrix no less than $I^{-1}$.

### 3.5 Inference

We note that the Fisher Information matrix $I$ does not have a closed form. A plug-in variance estimate as in Huang (1996) is not feasible. As an alternative, we propose using the weighted bootstrap developed in Ma and Kosorok (2005b).

Denote $w_1, w_2, \ldots, w_n$ as $n$ iid positive random weights generated from a known distribution with $\mathrm{P}(w) = 1$ and $var(w) = 1$. Denote $(\hat{\alpha}^*, \hat{\beta}^*, \hat{h}^* \hat{\Lambda}^*)$ as the weighted PMLE

$$(\hat{\alpha}^*, \hat{\beta}^*, \hat{h}^*, \hat{\Lambda}^*) = \arg\max_{\alpha,\beta,h,\Lambda} \mathrm{P}_n\{w \times l(X; \alpha, \beta, h, \Lambda)\} - \lambda_n^2 J^2(h).$$

Then conditional on the observed data, $(\hat{\alpha}^* - \hat{\alpha}, \hat{\beta}^* - \hat{\beta})$ has the same asymptotic variance as $(\hat{\alpha} - \alpha_0, \hat{\beta} - \beta_0)$.

Validity of the weighted bootstrap can be proved using Theorem 2 of Ma and Kosorok (2005b). For practical data analysis, we first generate $n$ i.i.d. weights from a known distribution, e.g., exp(1). The weighted PMLE defined above is then computed. We repeat this procedure $B$ times. The sample variance of the weighted estimates can be used to estimate variance of the PMLE. Asymptotically, any choice of weights satisfying the mean and variance requirements will lead to the same variance estimate. We refer to Ma and Kosorok (2005b) for more discussions.

## 4 Numerical studies

### 4.1 Computational algorithm

For the partly linear additive risk model, finding the PMLE is a constraint maximization problem, subject to the boundedness, smoothness and zero-mean constraint for $h$, and the non-decreasing constraint for $\Lambda$.

To estimate the spline function $h$, we take the approach proposed in Xiang and Wahba (1997), which states that an estimate with the number of basis functions growing at least at the rate of $n^{1/5}$ can achieve the same asymptotic precision as the full space. For practical data analysis, we suggest at least 20 basis functions. In our study, we choose equally spaced knots and the B-spline basis functions. Once the basis functions are chosen, maximization over the nonparametric $h$ becomes a parametric maximization problem. Maximization over $\alpha$, $\beta$ and the regression coefficients

in $h$ can be achieved with the Newton–Raphson method or functions such as *optim* in R. Maximization over the non-decreasing baseline function $\Lambda$ is achieved with the pool-adjacent-violator (PAV) approach.

The proposed PMLE involves the tuning parameter $\lambda_n$. Asymptotically, we assume that $\lambda_n$ satisfies A6. For practical data analysis, we set $\lambda_n = \tau \times n^{-1/3}$, where $\tau = 0.1, 0.2, \ldots, 5.0$. We propose using the five-fold cross validation (Wahba 1990) and searching over $\tau$ to determine its optimal value and the optimal value of $\lambda_n$.

## 4.2 Simulation study

We conduct simulation studies to investigate finite sample performance of the proposed estimate. The linear additive risk model is also considered, since it is simpler and has been extensively used. We consider the following four models:

*Model 1*. $\alpha_0 = (0, 1, 1)$ and $\beta_0 = (1, 1)$; With probabilities 0.5, $Z_1 = 0$ or 1.5; $Z_2 \sim$ *Unif* [0.2, 1.0]. The censoring time is generated independent of the covariates, exp(0.5) distributed and truncated with an upper bound of 7. The probability of cure is about 0.18. For subjects not cured, the censoring rate is about 0.27.

*Model 2*. $\alpha_0 = (0, 1, 1)$ and $\beta_0 = (1, 1)$; $Z_1 \sim$ *Unif* [0.5, 1.5] and $Z_2 \sim$ *Unif* [0.2, 1.0]. Censoring distribution is the same as in Model 1. The probability of cure is about 0.18. For subjects not cured, the censoring rate is about 0.27.

*Model 3*. $\alpha_0 = (0, 1, 1)$ and $\beta_0 = 1$. With probabilities 0.5, $Z_1 = 0.5$ or 1.5; $Z_2 \sim$ *Unif* $[0, 4]$. $h_0(Z_2) = sin(\pi Z_2)$. Censoring distribution is the same as in Model 1. The probability of cure is about 0.08. For subjects not cured, the censoring rate is about 0.24.

*Model 4*. $\alpha_0 = (0, 1, 1)$ and $\beta_0 = 1$. $Z_1 \sim$ *Unif* [0.5, 1.5] and $Z_2 \sim$ *Unif* $[0, 4]$. $h_0(Z_2) = sin(\pi Z_2)$. Censoring distribution is the same as in Model 1. The probability of cure is about 0.08. For subjects not cured, the censoring rate is about 0.24.

Models 1 and 2 are linear additive risk models, whereas Models 3 and 4 are partly linear additive risk models. In Models 1 and 3, one covariate has a discrete distribution, whereas in Models 2 and 4, both covariates are continuously distributed.

Summary statistics based on 500 replicates are shown in Table 1. We can see that the estimates have very small biases, especially with sample size 800. The standard deviations of the estimates shrink as the sample size increases from 400 to 800, at approximately the $\sqrt{2}$ rate. The weighted bootstrap estimates of standard errors are very close to the estimates' standard deviations.

In Fig. 1, we show the simulation results for Model 4 with sample size 400. The top panels show the histograms of estimated $\alpha$ (the second component) and $\beta$. We can see that the estimates have distributions very close to Normal. In the bottom panels, we show the point-wise means of estimated $h$ and $\Lambda$ and the 95% confidence intervals. We note that the point-wise confidence intervals are not formal inference. They are provided simply to show the variations of the estimates. We can see that the mean $\hat{h}$ matches the true $h_0$ very well, although the variations can be large. Estimated $h$ is less

**Table 1** Simulation study: summary statistics based on 500 replicates

| | | $\alpha_1$ | $\alpha_2$ | $\beta_1$ | $\beta_2$ |
|---|---|---|---|---|---|
| *Model 1* | | | | | |
| $n = 400$ | Est. | 0.813 | 0.825 | 1.080 | 1.194 |
| | SD | 0.630 | 1.085 | 0.391 | 0.670 |
| | $\hat{SD}$ | 0.654 | 0.986 | 0.412 | 0.713 |
| $n = 800$ | Est. | 0.938 | 0.973 | 1.037 | 1.098 |
| | SD | 0.510 | 0.881 | 0.258 | 0.449 |
| | $\hat{SD}$ | 0.430 | 0.826 | 0.279 | 0.484 |
| *Model 2* | | | | | |
| $n = 400$ | Est. | 0.736 | 0.896 | 1.139 | 1.171 |
| | SD | 0.874 | 1.447 | 0.687 | 1.202 |
| | $\hat{SD}$ | 0.959 | 1.502 | 0.655 | 1.026 |
| $n = 800$ | Est. | 0.803 | 0.944 | 1.199 | 1.012 |
| | SD | 0.666 | 1.113 | 0.511 | 0.812 |
| | $\hat{SD}$ | 0.660 | 1.058 | 0.583 | 0.728 |
| *Model 3* | | | | | |
| $n = 400$ | Est. | 1.007 | 1.021 | 1.102 | – |
| | SD | 0.458 | 0.409 | 0.418 | – |
| | $\hat{SD}$ | 0.412 | 0.413 | 0.400 | – |
| $n = 800$ | Est. | 0.989 | 1.058 | 1.070 | – |
| | SD | 0.273 | 0.292 | 0.329 | – |
| | $\hat{SD}$ | 0.261 | 0.320 | 0.364 | – |
| *Model 4* | | | | | |
| $n = 400$ | Est. | 0.938 | 1.027 | 1.161 | – |
| | SD | 0.574 | 0.469 | 0.519 | – |
| | $\hat{SD}$ | 0.558 | 0.462 | 0.374 | – |
| $n = 800$ | Est. | 1.007 | 1.020 | 0.996 | – |
| | SD | 0.384 | 0.311 | 0.333 | – |
| | $\hat{SD}$ | 0.412 | 0.321 | 0.360 | – |

*n* sample size; *Est.* mean of estimates; *SD* standard deviations of $\hat{\alpha}$ and $\hat{\beta}$; $\hat{SD}$ standard deviations of $\hat{\alpha}^* - \hat{\alpha}$ and $\hat{\beta}^* - \hat{\beta}$

satisfactory when it is close to the boundaries and there are fewer data points. Estimate of $\Lambda$ is also very satisfactory.

### 4.3 Calcification study

The Calcification study (Yu et al. 2001) investigated the calcification associated with hydrogel intraocular lenses, which is an infrequently reported complication of cataract treatment. The quantity of interest is the effect of clinical risk factors on the time to

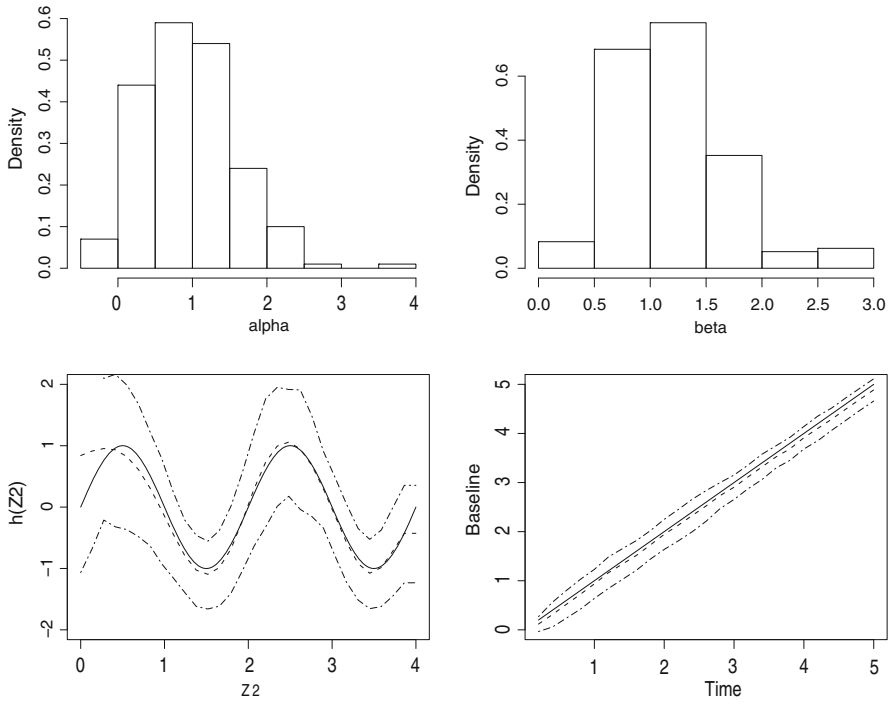


**Fig. 1** Simulation study (Model 4, sample size 400). *Upper left* histogram of the second component of $\hat{\alpha}$; *upper right* histogram of $\hat{\beta}$; *lower left* estimate of $h_0$, mean estimate and point-wise 95% confidence interval; *lower right* estimate of $\Lambda_0$, mean estimate and point-wise 95% confidence interval. The *solid lines* are the true values

calcification of the lenses after implantation. The patients were examined by an ophthalmologist to determine the status of calcification at a random time ranging from 0 to 36 months after implantation of the intraocular lenses. Current status data arises since only the examination time and the calcification status at examination are available. The longest follow-up was 3 years, but no new case was observed after 2 years. As has been noted by Yu et al. (2001), there is not enough evidence to conclude that the unaffected intraocular lenses will remain calcification free after 2 years. However it is highly likely that some patients are subject to much less risk of calcification, i.e., those patients may consist of a cured subgroup. The severity of calcification was graded on a discrete scale ranging from 0 to 4. Those with severity $\leq 1$ were classified as "not calcified". The clinical risk factors of interest include gender, incision length and age at implantation. The dataset contains 379 records. We exclude the one record with missing measurement, resulting in $n = 378$. For more discussions of the experimental setup, see Yu et al. (2001) and Lam and Xue (2005).

Let $Z_1 =$ incision length, $Z_2 =$ gender and $Z_3 =$ age at implantation/10. Denote $Z = (Z_1, Z_2, Z_3)$. We assume the logistic model for cure

$$\text{prob(not cure)} = \frac{\exp(\alpha_1 + \alpha_2 Z_3)}{1 + \exp(\alpha_1 + \alpha_2 Z_3)}. \tag{11}$$

For subjects susceptible to calcification, the conditional hazard satisfies the linear additive risk model

$$\lambda(T|Z) = \lambda(T) - (\beta_1 Z_1 + \beta_2 Z_2 + \beta_3 Z_3) \tag{12}$$

or the partly linear additive risk model

$$\lambda(T|Z) = \lambda(T) - (\beta_1 Z_1 + \beta_2 Z_2 + h(Z_3)). \tag{13}$$

The cure model (11) and the form of the covariate effects have been motivated by Lam and Xue (2005). We employ the proposed penalized approach, and select the tuning parameter using the five-fold cross validation. Inference is based on the weighted bootstrap with 500 realizations of random exp(1) weights. For model (12), we set $\lambda_n = 0$, and the PMLE becomes the MLE.

For model (12), the MLEs are

$$\hat{\alpha}_1 = -0.310(1.567); \quad \hat{\alpha}_2 = 1.225(0.332); \quad \hat{\beta}_1 = 0.022(0.036);$$
$$\hat{\beta}_2 = -0.047(0.044); \quad \hat{\beta}_3 = -0.051(0.009),$$

where values in "()" are the corresponding bootstrap standard deviation estimates. We can see that based on model (12), age has a significant effect on the cure probability: the cure rate decreases as the age increases. Out of the three covariates considered for survival, only the age effect is significant: older people have higher survival risks, which is intuitively reasonable.

We also consider model (13), the model that allows for nonparametric age effect in the additive risk model. The PMLEs are

$$\hat{\alpha}_1 = 1.686(1.462); \quad \hat{\alpha}_2 = 0.795(0.807); \quad \hat{\beta}_1 = 0.028(0.058);$$
$$\hat{\beta}_2 = -0.042(0.100).$$

We can see that estimates from model (13) are considerably different from their counterparts from model (12). The age effect on the cure rate is no longer significant. Incision length and gender still have no significant effect on survival. The age effect on survival has a bell shape (Fig. 2). We also show the estimated cumulative baseline hazard function with its lowess smoother in Fig. 2. The cumulative baseline function is almost linear, which suggests a constant hazard function.

In models (11) and (12), we have assumed specific forms of the covariate effects. Our choices have been motivated by Lam and Xue (2005). We note that such choices may be subject to more rigorous model checking. Since tools for formal model checking are still not available, interpretation of the data analysis results should be with extreme caution.

The same data has been analyzed in Lam and Xue (2005) and Ma (2007). In Lam and Xue (2005), the event time is modeled directly, and a straightforward estimation of the hazard function is not available. The Cox model in Ma (2007) also suggests no significant covariate effect on the cure rate. However, with the Cox model, the

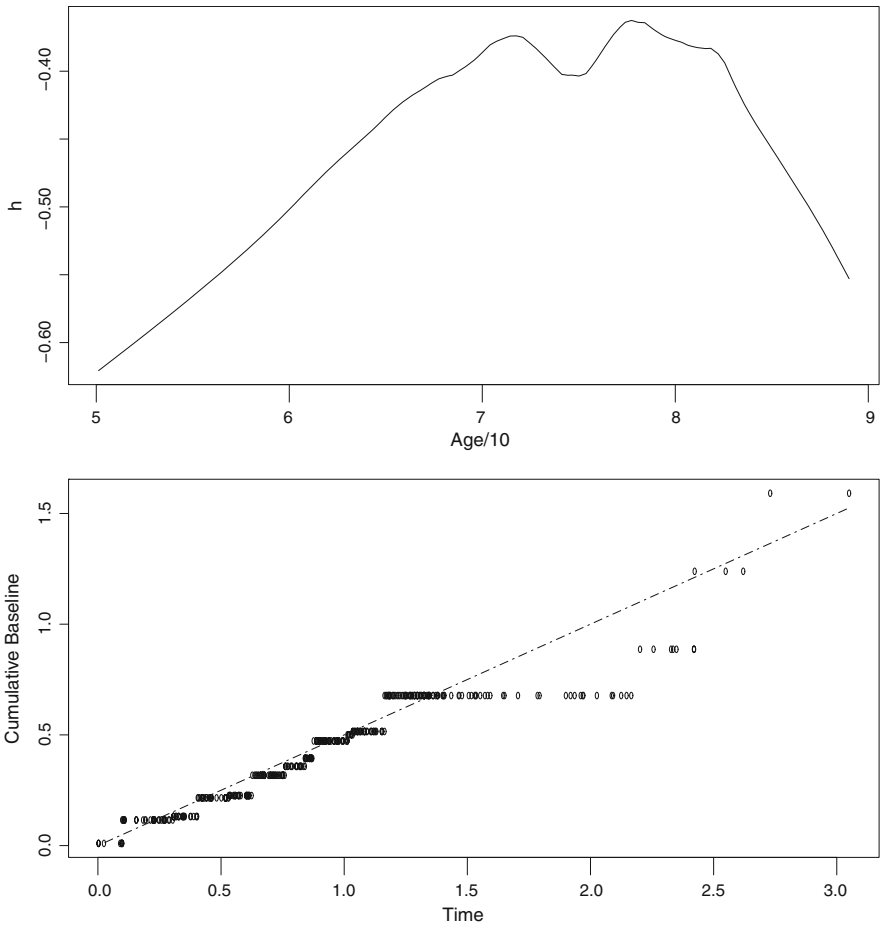


**Fig. 2** Calcification data under the partly linear model. Estimates of $h_0$ and $\Lambda_0$ (with lowess smoothers)

gender and incision length effects on survival are significant. The overall trends of the smooth age effects are similar under the Cox and additive models. We note that, without formal testing, we cannot conclude superiority of one model over the others. However, the additive model does provide an alternative way of interpreting the data. Specifically, different covariates contribute to the hazard in an additive manner, which makes the additive model the only possible one to investigate the "proportion" of hazard contributed by a specific covariate.

## 5 Discussions

In this article, we investigate the additive risk model for current status data with a cured subgroup. We propose the PMLE for estimation and the weighted bootstrap for inference. We show that, although the cumulative baseline hazard function and the nonparametric covariate effect can only be estimated at the $n^{1/3}$ convergence rate, the regression parameter is still $\sqrt{n}$ estimable, asymptotically normal and efficient. We

consider a full likelihood based approach. The proposed estimate is more efficient than the one in Lin (1998), and does not need to make assumptions on $\hat{\Lambda}$ as in Martinussen and Scheike (2002).

For simplicity of notations, we consider only one nonparametric covariate effect $h(Z_2)$. When there exist more nonparametric covariate effects, say $h_2(Z_2) + \cdots + h_M(Z_M)$, the proposed approach can be modified and employed. Specifically, the penalized estimate will become the double or multiple penalized estimate, with one penalty on each of the nonparametric covariate effect. The information calculation may become more complicated. However, the non-orthogonal projection approach can still be applied. Projection related to the nonparametric covariate effects can be carried out following Huang (1999). Lemma 4 and the validity of the weighted bootstrap will still hold. Computation can be carried out in a similar manner as in described in Sect. 4.1.

With the proposed two part modeling, we consider a mixture of the survival function, i.e, $S = p + (1 - p) \times \tilde{S}$, where $\tilde{S}$ is a proper survival function and $p > 0$. In principal, it is possible to consider other mixture models. For example, as has been pointed out by one of the reviewers, we may consider $\lambda = \gamma \times \tilde{\lambda}$, where $\tilde{\lambda}$ is an ordinary hazard function and $\gamma$ is the frailty term with point mass at zero. Then the hazard function will be a mixture of zero and an ordinary hazard. In theory, other type of mixture models can be investigated in a similar manner as presented in this article. However, such extensions can be highly non-trivial and will not be pursued in this article.

Model diagnosis is not investigated in this study. We suspect that previously proposed techniques based on cumulative residuals may be applicable. However, we fail to prove theoretically their applicability and decide not to use them. Another relevant unsolved problem is how to compare the proposed additive risk model with the AFT model or the Cox model under present data settings. Due to existence of cured subjects and interval censoring, standard model comparison tools, such as the time-dependent ROC for right censored data, cannot be utilized here. Prediction of survival risks based method is being developed for comparison purpose and will be reported in a separate study. Our study is under the assumption that the cured subgroup exists, which may or may not be true. Thus, another possible extension of our study is to develop methods for detecting existence of the cured subgroup with interval censored data, which needs uniform convergence of $\hat{\Lambda}$ over a finite interval. Our limited investigation shows that this is an extremely hard problem and will not be pursued here.

**Acknowledgments** The author would like to thank the editor and the referees for insightful comments, that have led to significant improvement of the paper. This study is partly supported by grant #0805984 from NSF, USA.

## Appendix

*Proof of Lemma 2*

**Definition** (*Bracketing number*). Let $(\mathbb{F}, ||\cdot||)$ be a subset of a normed space of real function $f$ on some set. Given two functions $f_1$ and $f_2$, the bracket $[f_1, f_2]$ is the set of all functions $f$ with $f_1 \leq f \leq f_2$. An $\epsilon$ bracket is a bracket $[f_1, f_2]$ with $||f_1 - f_2|| \leq \epsilon$. The bracketing number $N_{[]}(\epsilon, \mathbb{F}, ||\cdot||)$ is the minimum number

of $\epsilon$ brackets needed to cover $\mathbb{F}$. The entropy with bracketing is the logarithm of the bracketing number.

Lemma 25.84 of van der Vaart (1998) shows that there exists a constant $K_2$ such that for every $\epsilon > 0$, $\log N_{[]}(\epsilon, \{\Lambda\}, L_2(P)) \leq K_2(\frac{1}{\epsilon})$, if assumption A3 is satisfied.

van de Geer (2000) shows that for the class

$$\tilde{\mathbb{H}} = \left\{ h : [0,1] \to [0,1] \int (h^{(s_0)}(x))^2 dx < 1 \right\},$$

we have $\log N_{[]}(\epsilon, \tilde{\mathbb{H}}, L_2(P)) \leq K_3 \epsilon^{-1/s_0}$, for a fixed constant $K_3$, $s_0 \geq 1$ and all $\epsilon$. This result, combined with the entropy calculation for $\{\Lambda\}$, give that

$$\log N_{[]}(\epsilon, \{l(\alpha, \beta, h, \Lambda)\}, L_2(P)) \leq K_4 \epsilon^{-1}, \tag{14}$$

for a fixed constant $K_4$.

From the definition of the PMLE, we have

$$\mathrm{P}_n l(\hat{\alpha}, \hat{\beta}, \hat{h}, \hat{\Lambda}) - \lambda_n^2 J^2(\hat{h}) \geq \mathrm{P}_n l(\alpha_0, \beta_0, h_0, \Lambda_0) - \lambda_n^2 J^2(h_0), \tag{15}$$

which can also be written as

$$\begin{aligned} &\lambda_n^2 J^2(\hat{h}) + \mathrm{P}[l(\alpha_0, \beta_0, h_0, \Lambda_0) - l(\hat{\alpha}, \hat{\beta}, \hat{h}, \hat{\Lambda})] \\ &\leq \lambda_n^2 J^2(h_0) + (\mathrm{P}_n - \mathrm{P})[l(\alpha_0, \beta_0, h_0, \Lambda_0) - l(\hat{\alpha}, \hat{\beta}, \hat{h}, \hat{\Lambda})] \end{aligned} \tag{16}$$

Applying the entropy result in (14), we have

$$(\mathrm{P}_n - \mathrm{P})[l(\alpha_0, \beta_0, h_0, \Lambda_0) - l(\hat{\alpha}, \hat{\beta}, \hat{h}, \hat{\Lambda})] = (1 + J(h_0) + J(\hat{h}))o_p(n^{-1/2}). \tag{17}$$

Combine inequalities (16) and (17). Simple calculations show that $\lambda_n J(\hat{h}) = o_p(1)$.

Equation (16) and assumption A5 hence yield

$$K_1 d^2((\hat{\alpha}, \hat{\beta}, \hat{h}, \hat{\Lambda}), (\alpha_0, \beta_0, h_0, \Lambda_0)) \leq o_p(1) + (1 + J(h_0) + J(\hat{h}))o_p(n^{-1/2}).$$

We can conclude that the PMLE is consistent.

To prove the rate of convergence, we use the following result.

(Theorem in van de Geer 2000, Page 79). Consider a uniformly bounded class of functions $\Gamma$, with $\sup_{\gamma \in \Gamma} |\gamma - \gamma_0|_\infty < \infty$, a fixed $\gamma_0 \in \Gamma$, and $\log N_{[]}(\epsilon, \Gamma, P) \leq K_9 \epsilon^{-b}$ for all $\epsilon > 0$. Here $b \in (0, 2)$ and $K_9$ is a fixed constant. Then for $\delta_n = n^{-1/(2+b)}$,

$$\sup_{\gamma \in \Gamma} \frac{|(\mathrm{P}_n - \mathrm{P})(\gamma - \gamma_0)|}{||\gamma - \gamma_0||_2^{1-b/2} \vee \sqrt{n}\delta_n^2} = O_p(n^{-1/2}), \tag{18}$$

where $x \vee y = \max(x, y)$.

Considering the compactness assumptions A1–A3, uniqueness assumption A5 and the smoothness of the objective function, we have

$$K_1 d^2((\hat{\alpha}, \hat{\beta}, \hat{h}, \hat{\Lambda}), (\alpha_0, \beta_0, h_0, \Lambda_0)) \leq \mathrm{P}[l(\alpha_0, \beta_0, h_0, \Lambda_0) - l(\hat{\alpha}, \hat{\beta}, \hat{h}, \hat{\Lambda})] \leq K_5 d^2((\hat{\alpha}, \hat{\beta}, \hat{h}, \hat{\Lambda}), (\alpha_0, \beta_0, h_0, \Lambda_0)), \tag{19}$$

for a fixed constant $K_5$.

Combining (16) with (18) for $b = 1$ and (19), we have

$$\begin{aligned} &\lambda_n^2 J^2(\hat{h}) + K_1 d^2((\hat{\alpha}, \hat{\beta}, \hat{h}, \hat{\Lambda}), (\alpha_0, \beta_0, h_0, \Lambda_0)) \\ &\leq \lambda_n^2 J^2(h_0) + (1 + J(h_0) + J(\hat{h}))O_p(n^{-1/2}) \\ &\quad \times (d^{1/2}((\hat{\alpha}, \hat{\beta}, \hat{h}, \hat{\Lambda}), (\alpha_0, \beta_0, h_0, \Lambda_0)) \vee n^{-1/6}). \end{aligned} \tag{20}$$

We conclude from (20) that

$$\begin{aligned} \lambda_n^2 J^2(\hat{h}) &\leq \lambda_n^2 J^2(h_0) + (1 + J(h_0) + J(\hat{h}))O_p(n^{-1/2}) \\ &\quad \times (d^{1/2}((\hat{\alpha}, \hat{\beta}, \hat{h}, \hat{\Lambda}), (\alpha_0, \beta_0, h_0, \Lambda_0)) \vee n^{-1/6}), \\ K_1 d^2((\hat{\alpha}, \hat{\beta}, \hat{h}, \hat{\Lambda}), (\alpha_0, \beta_0, h_0, \Lambda_0)) &\leq \lambda_n^2 J^2(h_0) + (1 + J(h_0) + J(\hat{h})) \\ &\quad \times O_p(n^{-1/2})(d^{1/2}((\hat{\alpha}, \hat{\beta}, \hat{h}, \hat{\Lambda}), (\alpha_0, \beta_0, h_0, \Lambda_0)) \vee n^{-1/6}). \end{aligned}$$

Simple calculations give that

$$J(\hat{h}) = O_p(1) \quad \text{and} \quad d((\hat{\alpha}, \hat{\beta}, \hat{h}, \hat{\Lambda}), (\alpha_0, \beta_0, h_0, \Lambda_0)) = O_p(n^{-1/3}). \qquad \square$$

*Proof of Lemma 4* We recall that $\dot{l}_{\alpha\beta} = (\dot{l}_\alpha, \dot{l}_\beta)'$. Define $\dot{l}_{h\Lambda}[\xi, a] = (\dot{l}_h(\xi), \dot{l}_\Lambda(a))'$. From assumption A6 and Lemma 2, the PMLE "nearly" maximizes the likelihood function, i.e,

$$\mathrm{P}_n \dot{l}_{\alpha\beta}(\hat{\alpha}, \hat{\beta}, \hat{h}, \hat{\Lambda}) = o_P(n^{-1/2}), \quad \mathrm{P}_n \dot{l}_{h\Lambda}[\xi, a](\hat{\alpha}, \hat{\beta}, \hat{h}, \hat{\Lambda}) = o_P(n^{-1/2}),$$

for any $\xi \in \mathbb{B}$ and $a \in \mathbb{A}$. In the above equations, $o_P$ should be understood as component-wise. In addition, we have:

1. *Consistency and convergence rate*. $|\hat{\alpha} - \alpha_0| = O_p(n^{-1/3})$, $|\hat{\beta} - \beta_0| = O_p(n^{-1/3})$, $||\hat{h} - h_0||_2 = O_p(n^{-1/3})$ and $||\hat{\Lambda} - \Lambda_0||_2 = O_p(n^{-1/3})$ from Lemma 2.
2. *Positive information* The Fisher Information matrix is positive definite and component wise bounded from Lemma 3.
3. *Stochastic equicontinuity*. For any $\delta_n \to 0$ and constant $K_6 > 0$, within the neighborhood $\{|\alpha - \alpha_0| < \delta_n, |\beta - \beta_0| < \delta_n, ||h - h_0||_2 < K_6 n^{-1/3}, ||\Lambda - \Lambda_0||_2 < K_6 n^{-1/3}\}$,

$$\begin{aligned} &\sup \sqrt{n} |(\mathrm{P}_n - \mathrm{P})(\dot{l}_{\alpha\beta}(\alpha, \beta, h, \Lambda) - \dot{l}_{\alpha\beta}(\alpha_0, \beta_0, h_0, \Lambda_0))| = o_p(1), \\ &\sup \sqrt{n} |(\mathrm{P}_n - \mathrm{P})(\dot{l}_{h\Lambda}[\xi, a](\alpha, \beta, h, \Lambda) - \dot{l}_{h\Lambda}[\xi^*, m^*](\alpha_0, \beta_0, h_0, \Lambda_0))| = o_p(1), \end{aligned}$$

where $a \in \mathbb{A}, \xi \in \mathbb{B}, \xi^*$ is as defined in Lemma 3 and $m^* = \mathrm{P}(\dot{l}_{\alpha\beta}\tilde{l}_\Lambda|C)/\mathrm{P}(\tilde{l}_\Lambda\tilde{l}_\Lambda|C) - \mathrm{P}(\dot{l}_h(\xi^*)\tilde{l}_\Lambda|C)/\mathrm{P}(\tilde{l}_\Lambda\tilde{l}_\Lambda|C)$.

4. *Smoothness of the model*. For $(\alpha, \beta, h, \Lambda)$ within the neighborhood $\{|\alpha - \alpha_0| < \delta_n, |\beta - \beta_0| < \delta_n, ||h - h_0||_2 < K_6 n^{-1/3}, ||\Lambda - \Lambda_0||_2 < K_6 n^{-1/3}\}$, the expectations of $\dot{l}_{\alpha\beta}$ and $\dot{l}_{h\Lambda}$ are Hellinger differentiable.

Conditions in Theorem 3.4 of Huang (1996) are satisfied. The asymptotical normality and efficiency thus follow. □

## References


Banerjee, S., Carlin, B. P. (2004). Parametric spatial cure rate models for interval-censored time-to-relapse data. *Biometrics, 60*, 268–275.

Bellamy, S. L., Li, Y., Ryan, L. M., Lipsitz, S., Canner, M. J., Wright, R. (2004). Analysis of clustered and interval censored data from a community-based study in asthma. *Statistics in Medicine, 23*, 3607–3621.

Chen, M., Ibrahim, J. G., Sinha, D. (2004). A new joint model for longitudinal and survival data with a cure fraction. *Journal of Multivariate Analysis, 91*, 18–34.

Diamond, I. D., McDonald, J. W. (1991). Analysis of current status data. In *Demographic applications of event history analysis* (pp. 231–252). NY, USA: Oxford University Press.

Farewell, V. T. (1986). Mixture models in survival analysis: Are they worth the risk? *Canadian Journal of Statistics, 14*, 257–262.

Gart, J. J., Krewski, D., Lee, P. N., Tarone, R. E., Wahrendorf, J. (1986). *Statistical methods in cancer research: Vol. III. The design and analysis of longterm animal experiments*. IARC Scientific Publications No. 79. Lyon: International Agency for Research on Cancer.

Groeneboom, P., Wellner, J. A. (1992). *Information bounds and nonparametric maximum likelihood estimation*. Basel: Birkhauser.

Huang, J. (1996). Efficient estimation for the proportional hazard model with interval censoring. *The Annals of Statistics, 24*, 540–568.

Huang, J. (1999). Efficient estimation of the partly linear additive Cox model. *Annals of Statistics, 27*, 1536–1563.

Lam, K. F., Xue, H. (2005). A semiparametric regression cure model with current status data. *Biometrika, 92*, 573–586.

Li, C. S., Taylor, J. M. G., Sy, J. P. (2001). Identifiability of cure models. *Statistics and Probability Letters, 54*, 389–395.

Lin, D. Y., Oakes, D., Ying, Z. (1998). Additive hazards regression with current status data. *Biometrika, 85*, 289–298.

Lu, W., Ying, Z. L. (2004). On semiparametric transformation cure model. *Biometrika, 91*, 331–343.

Ma, S. (2007). Cure model with current status data. *Statistica Sinica* (in press).

Ma, S., Kosorok, M. R. (2005a). Penalized log-likelihood estimation for partly linear transformation models with current status data. *Annals of Statistics, 33*, 2256–2290.

Ma, S., Kosorok, M. R. (2005b). Robust semiparametric M-estimation and the weighted bootstrap. *Journal of Multivariate Analysis, 96*, 190–217.

Martinussen, T., Scheike, T. H. (2002). Efficient estimation in additive hazards regression with current status data. *Biometrika, 89*, 649–658.

Peng, Y., Dear, K. B. G. (2000). A nonparametric mixture model for cure rate estimation. *Biometrics, 56*, 237–243.

Sasieni, P. (1992). Non-orthogonal projections and their application to calculating the information in a partly linear Cox model. *Scandinavian Journal of Statistics, 19*, 215–233.

Shiboski, S. C., Jewell, N. P. (1992). Statistical analysis of the time dependence of HIV infectivity based on partner study data. *Journal of the American Statistical Association, 87*, 360–372.

Thompson, L. A., Chhikara, R. S. (2003). A Bayesian cure rate model for repeated measurements and interval censoring. In *Proceedings of the joint statistical meeting 2003*.

van de Geer, S. (2000). *Empirical processes in M-Estimation*. Cambridge Series in Statistical and Probabilistic Mathematics.

van der Vaart, A. W. (1998). *Asymptotic statistics*. Cambridge: Cambridge University Press.

Wahba, G. (1990). *Spline models for observational data.* SIAM. CBMS-NSF Regional Conference Series in Applied Mathematics.

Xiang, D., Wahba, G. (1997). Approximate smoothing spline methods for large data set in the binary case. In *Proceedings of the joint statistical meetings*. Biometrics Section, pp. 94–98.

Xue, H., Lam, K. F., Li, G. (2004). Sieve maximum likelihood estimator for semiparametric regression models with current status data. *Journal of the American Statistical Association, 99*, 346–356.

Yu, A. K. F., Kwan, K. Y. W., Chan, D. H. Y., Fong, D. Y. T. (2001). Clinical features of 46 eyes with calcified hydrogel intraocular lenses. *Journal of Cataract and Refractive Surgery, 27*, 1596–1606.