CURE MODEL WITH CURRENT STATUS DATA

Shuangge Ma

Yale University

Abstract: Current status data arise when only random censoring time and event status at censoring are observable. We consider current status data under the cure model, where a proportion of the subjects are not susceptible to the event of interest. We assume a generalized linear model for the cure probability. For subjects not cured, the linear and partly linear Cox proportional hazards models are used to model the survival risk. We propose estimation using the (penalized) maximum likelihood approach. It is shown that estimates of the parametric regression coefficients are $\sqrt{n}$ consistent, asymptotically normal and efficient. The nonparametric baseline function and nonparametric covariate effect can be estimated with $n^{1/3}$ convergence rate. We propose inference for estimates of the regression coefficients using the weighted bootstrap. Simulation studies are used to assess finite sample performance of the proposed estimates. We analyze the Calcification data using the proposed approach.

Key words and phrases: Cure model, current status data, M-estimator.

1. Introduction

Current status data arise naturally in medical studies where the target measurement is the time of event occurrence, but observations are limited to indicators of whether or not the event has occurred at the time the sample is collected, i.e., only the current status of each observation with respect to event occurrence is observable. Examples of current status data include the demographic study of age at weaning (Grummer-Strawn (1993)), clinical study of tumor occurrence (Gart, Krewski, Lee, Tarone and Wahrendorf (1986)) and HIV transmission among sexually partners (Jewell and Shiboski (1990)), among many others. Previous methodological studies include the nonparametric model in Groeneboom and Wellner (1992), the linear Cox model in Huang (1996), the additive risk model in Lin, Oakes and Ying (1998), the partly linear accelerated failure time (AFT) model in Xue, Lam and Li (2004), and the partly linear transformation model in Ma and Kosorok (2005a). In the aforementioned studies, it has been assumed that if the followup time is long enough, all subjects will eventually experience the event of interest.

In this article, we consider current status data with a cured subgroup, where individuals in that subgroup are not susceptible to the event. Our research is motivated by the Calcification study of hydrogel intraocular lenses, which is an infrequently reported complication of cataract treatment (Yu, Kwan, Chan and Fong (2001)). In the Calcification study, 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 arise since only the examination time and the calcification status at examination are available. The longest follow-up was three years, but no new case was observed after two years. It is pointed out by Yu et al. (2001) that there is not enough evidence to conclude that the unaffected intraocular lenses will remain calcification-free after two years. However it is highly likely that some subjects are subject to much less risk of calcification, i.e., those subjects may consist of a cured sub-population.

For right censored survival data, studies of cure models include Farewell (1986), Kuk and Chen (1992), Lu and Ying (2004), Peng and Dear (2000), Li, Taylor and Sy (2001), Sy and Taylor (2000, 2001), Taylor (1995), Chen, Ibrahim and Sinha (2004) and Fang, Li and Sun (2005), among others. For interval censored data with a cured subgroup, Lam and Xue (2005) assume a partly linear AFT (accelerated failure time) model for susceptible subjects and consider a sieve maximum likelihood approach; Thompson and Chhikara (2003) propose a parametric model for the event time and a Bayesian estimator; when correlation exists among subjects, a parametric model for the event time is investigated in Bellamy, Li, Ryan, Lipsitz, Canner and Wright (2004); when spatial association exists among subjects, a frailty model with a Bayesian estimator is proposed in Banerjee and Carlin (2004).

For 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 linear or partly linear Cox models (Cox (1972)). We propose using (penalized) maximum likelihood approach for estimation, and the weighted bootstrap for inference. The assumed models are more flexible than the parametric models in Bellamy et al. (2004) for example . Compared with Lam and Xue (2005), the more popular Cox model is assumed, which is theoretically more challenging due to presence of the nonparametric baseline function. For the partly linear Cox model, we consider a penalized approach, which provides a flexible alternative to the sieve approach in Lam and Xue (2005). Although the empirical process techniques used to establish asymptotic properties of the proposed estimates have been mostly developed in van der Vaart and Wellner (1996), van de Geer (2000), Huang (1996) and

Ma and Kosorok (2005a,b), implementing them to study the proposed models is far from trivial. The proposed models, together with their theoretical properties, provide more insights into interval censored survival study when there exists a cured subgroup.

We introduce the data structure in Section 2. Linear and partly linear Cox models are investigated in Sections 3 and 4, respectively. Asymptotic properties of the maximum likelihood estimate (MLE) and penalized maximum likelihood estimate (PMLE) are established. Numerical studies, including simulations and analysis of the Calcification data, are presented in Section 5. The article concludes with a discussion in Section 6. Proofs are provided in the Appendix.

2. Data and Model Settings

Let T be the event time of interest, and C the random censoring. For simplicity, we assume only two covariates Z_1 and Z_2, with $Z = (Z_1, Z_2)'$. 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 ($T = \infty$), and $U = 1$ otherwise.

We model the cure probability using a generalized linear model with a known link function. We are especially interested in the logistic model

$$P(U = 1|Z) = \phi(\alpha'\tilde{Z}) = \frac{\exp(\alpha'\tilde{Z})}{1 + \exp(\alpha'\tilde{Z})}. \qquad (2.1)$$

Here α is the unknown regression coefficient, α' is the transpose of α, and $\tilde{Z} = (1, Z')'$. Besides the logistic function, ϕ might be the identify function or the log function, among others.

For subjects with $U = 1$, we model the survival risk with the Cox proportional hazards model (Cox (1972)):

$$\Lambda(T|Z) = \Lambda(T)\exp(-f(Z)), \qquad (2.2)$$

where $\Lambda(T|Z)$ is the conditional cumulative hazard function, $\Lambda(T)$ is the unknown cumulative baseline function, and $f(Z)$ is the covariate effect. We write $\Lambda(T)$ as Λ hereafter. We consider the linear Cox model

$$f(Z) = \beta_1 Z_1 + \beta_2 Z_2, \qquad (2.3)$$

and the partly linear Cox model

$$f(Z) = \beta_1 Z_1 + h(Z_2), \qquad (2.4)$$

where β_1 and β_2 are unknown regression coefficients and h is an unknown smooth covariate effect. Further flexibility is introduced in (2.4) by allowing for nonparametric covariate effect h.

Assume n i.i.d. observations $X_1 = (C_1, \delta_1, Z_{11}, Z_{21}), \ldots, X_n = (C_n, \delta_n, Z_{1n}, Z_{2n})$ are available.

3. Linear Cox Model

Suppose event time and censoring are conditionally independent. Under the linear Cox model (2.3), the log-likelihood for a single observation is

$$l_1(X) = \delta \log(\phi(\alpha'\tilde{Z})) + \delta \log[1 - \exp(-\Lambda(C)\exp(-\beta'Z))]$$
$$+ (1 - \delta)\log\{1 - \phi(\alpha'\tilde{Z})[1 - \exp(-\Lambda(C)\exp(-\beta'Z))]\}, \qquad (3.1)$$

up to a constant. Let $\beta = (\beta_1, \beta_2)'$. Based on n i.i.d. observations, the maximum likelihood estimate (MLE) is

$$(\hat{\alpha}, \hat{\beta}, \hat{\Lambda}) = argmax_{\alpha,\beta,\Lambda} E_n l_1(X), \qquad (3.2)$$

where E_n is expectation under the empirical measure P_n.

Denote the true value of (α, β, Λ) by $(\alpha_0, \beta_0, \Lambda_0)$. We 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 (α_0, β_0) belongs to a compact subset of $\mathbb{R}^5$; (3) For any $\alpha \neq \alpha_0$, $Pr(\alpha'\tilde{Z} \neq \alpha_0'\tilde{Z}) > 0$, and for any $\beta \neq \beta_0$, $Pr(\beta'Z \neq \beta_0'Z) > 0$.

A3. (1) For $l_C \le T \le u_C$, there exists a constant M such that $0 < 1/M < \Lambda_0 < M < \infty$; (2) Λ_0 has strictly positive first order derivative on $[l_C, u_C]$.

Define the distance between (α, β, Λ) and $(\alpha_0, \beta_0, \Lambda_0)$ as

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

where $\|\Lambda(c) - \Lambda_0(c)\|_2^2 = \int_{l_C}^{u_C} (\Lambda(c) - \Lambda_0(c))^2 dc$. We also assume

A4. For (α, β, Λ) satisfying assumptions A1−A3,

$$E[l_1(\alpha, \beta, \Lambda) - l_1(\alpha_0, \beta_0, \Lambda_0)] \le -K_1 d^2((\alpha, \beta, \Lambda), (\alpha_0, \beta_0, \Lambda_0))$$

for a fixed constant $K_1 > 0$.

Similar assumptions have been made in Huang (1996) and van der Vaart (1998). We note that in the compactness assumptions, the actual bounds may remain unknown – they are not needed in the proof or for computation.

Remark 1. For subjects with $U = 1$, the survival function S should satisfy $S(\infty) = 0$. We note that under the compactness assumptions in A1-A3, $S(u_C) > 0$. This does not contradict the $S(\infty) = 0$ assumption, as S for $T > u_C$ cannot be estimated. For data analysis, we can set $\hat{\Lambda} = \infty$ for $T > max\{C_n\}$. Related practical issues have been discussed in Li et al. (2001).

Under A1−A4, the proposed model is identifiable. Proof follows from Li et al. (2001). In data analysis, a cure model should not be suggested unless a clear level plateau of the survival function is observed (Fang, Li and Sun (2005) and Taylor (1995)). When the estimated cure probabilities are very close to one or zero for most subjects, an identifiability problem may exist. In this case, models with and without the cure proportion assumption should be fitted and compared (Taylor (1995)).

Remark 2. Compactness assumptions are made in A1-A3 so that the MLE defined in (3.2) exists. We note that for any finite n, we do not assume the MLE is unique. However under A4, the MLE is asymptotically unique as $n \to \infty$.

3.1. Finite sample properties

Let $C_{(1)}, \ldots, C_{(n)}$ be the order statistics of $C_1, \ldots, C_n$. Let $\delta_{(i)}, Z_{(1i)}, Z_{(2i)}$ correspond to $C_{(i)}$. Since only the values of Λ at $C_{(i)}$ matter in the log-likelihood function, we take the maximum likelihood estimate $\hat{\Lambda}$ to be the right-continuous non-decreasing step function with jump points only at $C_{(i)}$.

Lemma 1. *The MLE defined in* (3.2) *satisfies*

$$\frac{\partial E_n l_1}{\partial \alpha}\Big|_{\alpha=\hat{\alpha},\beta=\hat{\beta},\Lambda=\hat{\Lambda}} = 0, \quad \frac{\partial E_n l_1}{\partial \beta}\Big|_{\alpha=\hat{\alpha},\beta=\hat{\beta},\Lambda=\hat{\Lambda}} = 0, \tag{3.3}$$

$$\sum_{j \geq i} \left(\frac{\delta_{(j)}}{1 - \exp(-\hat{\Lambda}_{(j)} \exp(-\hat{\beta}' Z_{(j)}))} - \frac{(1 - \delta_{(j)})\phi}{1 - \phi[1 - \exp(-\hat{\Lambda}_{(j)} \exp(-\hat{\beta}' Z_{(j)}))]} \right)$$
$$\times \exp(-\hat{\beta}' Z_{(j)}) \exp(-\hat{\Lambda}_{(j)} \exp(-\hat{\beta}' Z_{(j)})) \leq 0, \tag{3.4}$$

$$\sum_{i=1}^{n} \left(\frac{\delta_{(i)}}{1 - \exp(-\hat{\Lambda}_{(i)} \exp(-\hat{\beta}' Z_{(i)}))} - \frac{(1 - \delta_{(i)})\phi}{1 - \phi[1 - \exp(-\hat{\Lambda}_{(i)} \exp(-\hat{\beta}' Z_{(i)}))]} \right)$$
$$\times \exp(-\hat{\beta}' Z_{(i)}) \exp(-\hat{\Lambda}_{(i)} \exp(-\hat{\beta}' Z_{(i)}))\hat{\Lambda}_{(i)} = 0, \tag{3.5}$$

for $i = 1, \ldots, n$. Equations (3.3) hold simply from the definition of the MLE. Equations (3.4) and (3.5) can be proved by following Proposition 1.1 of Groeneboom and Wellner (1992). The proof is omitted here.

3.2. Consistency and rate of convergence

Lemma 2. (Consistency). *Under* A1$-$A4, $d((\hat{\alpha}, \hat{\beta}, \hat{\Lambda}), (\alpha_0, \beta_0, \Lambda_0)) = o_p(1)$.

The proof of Lemma 2 is in the Appendix. Based on the consistency result, we can establish the following convergence rate.

Lemma 3. (Convergence rate). *Under* A1$-$A4, $d((\hat{\alpha}, \hat{\beta}, \hat{\Lambda}), (\alpha_0, \beta_0, \Lambda_0)) = O_p(n^{-1/3})$.

Groeneboom and Wellner (1992) and Huang (1996) prove that, under assumptions similar to A1-A4, the best possible convergence rate for estimates of the nonparametric baseline function with current status data is $n^{1/3}$. This is considerably slower than the $n^{1/2}$ rate for right censored data, due to the excessive censoring. Lemma 3 shows that the optimal convergence rate can in fact be achieved under the proposed linear Cox model.

3.3. Fisher information

The score functions for α and β are the first order derivatives of the log-likelihood function:

$$\dot{l}_{1\alpha} = \dot{\phi}\left(\frac{\delta}{\phi} - \frac{(1-\delta)g_1}{1 - \phi g_1}\right)\tilde{Z},$$

$$\dot{l}_{1\beta} = -\left(\frac{\delta}{g_1} - \frac{(1-\delta)\phi}{1 - \phi g_1}\right)\Lambda \exp(-\beta' Z)\exp(-\Lambda\exp(-\beta' Z))Z,$$

where $\dot{\phi}$ is the derivative of ϕ and $g_1(\beta, \Lambda) = 1 - \exp(-\Lambda\exp(-\beta' Z))$. Let $\mathbb{A} = \{a : \int_{l_C}^{u_C} a(c)dc = 0;\ a \in L_2(P);$ for u small enough $\Lambda_u = \Lambda + ua$ satisfies $A3\}$. Then

$$\dot{l}_{1\Lambda}(a) = \left(\frac{\delta}{g_1} - \frac{(1-\delta)\phi}{1 - \phi g_1}\right)\exp(-\beta' Z)\exp(-\Lambda\exp(-\beta' Z))a = \tilde{l}_{1\Lambda}a.$$

Project the score functions of α and β onto the space generated by $\tilde{l}_{1\Lambda}a$. The efficient scores for α and β are

$$U_{1\alpha} = \dot{l}_{1\alpha} - \tilde{l}_{1\Lambda}\frac{\mathrm{P}(\dot{l}_{1\alpha}\tilde{l}_{1\Lambda}|C)}{\mathrm{P}(\tilde{l}_{1\Lambda}\tilde{l}_{1\Lambda}|C)} \quad\text{and}\quad U_{1\beta} = \dot{l}_{1\beta} - \tilde{l}_{1\Lambda}\frac{\mathrm{P}(\dot{l}_{1\beta}\tilde{l}_{1\Lambda}|C)}{\mathrm{P}(\tilde{l}_{1\Lambda}\tilde{l}_{1\Lambda}|C)}.$$

Write $I_1 = \mathrm{P}\{(U_{1\alpha}, U_{1\beta})'(U_{1\alpha}, U_{1\beta})\}$ and assume

A5. I_1 is positive definite and component-wise bounded.

Lemma 4. (Fisher Information). *Under* A1−A5, I_1 *is the efficient information matrix for* (α, β).

3.4. Asymptotic normality

Lemma 5. (Asymptotic normality). *Under* A1−A5,

$$\sqrt{n}(\hat{\alpha} - \alpha_0, \hat{\beta} - \beta_0) \to_d N(0, I_1^{-1}(\alpha_0, \beta_0, \Lambda_0)).$$

Lemma 5 indicates that the regression parameters are still $\sqrt{n}$ estimable, despite the slow convergence rate of the cumulative baseline estimate. Moreover, the proposed MLE is asymptotically efficient, in the sense that any regular estimator has asymptotic covariance matrix no less than I_1^{-1}.

3.5. Inference

Inference for $(\hat{\alpha}, \hat{\beta})$ can be based on the asymptotic normality result in Lemma 5 and a plug-in variance estimate. Numerical studies (omitted here) show that the plug-in estimate can be unreliable unless the sample size is very large. As an alternative, we consider the following weighted bootstrap, which is computationally intensive but may be preferred for data with small to medium sample sizes.

Lemma 6. (Weighted bootstrap). *Let* $w_1, \ldots, w_n$ *be* n *i.i.d. realizations of positive random weights generated from a known distribution with* $E(w) = 1$ *and* $var(w) = 1$. *The weighted MLE is*

$$(\hat{\alpha}^*, \hat{\beta}^*, \hat{\Lambda}^*) = \arg \max_{\alpha, \beta, \Lambda} \sum_i w_i l_1(X_i).$$

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

Lemma 6 is a straightforward application of Theorem 2 in Ma and Kosorok (2005b). For data analysis, we first generate B (for example $B = 500$) realizations of random weights. We propose using exp(1) distributed weights. For each realization of the random weights, the weighted MLE defined in Lemma 6 is computed. Repeat the weighted estimation B times. The sample variance of the weighted estimates provides an honest estimate of the variance of the MLE.

Remark 3. Any choice of weights satisfying the mean and variance requirements leads to asymptotically the same variance estimate. Very small differences are

observed for data analysis using different weights. It is not clear whether the ordinary bootstrap holds for the proposed model. We refer to Ma and Kosorok (2005b) for more detailed discussions.

4. Partly Linear Cox Model

Under the partly linear model, the log-likelihood for a single observation is

$$l_2(X) = \delta \log(\phi(\alpha'\tilde{Z})) + \delta \log[1 - \exp(-\Lambda \exp(-(\beta_1 Z_1 + h(Z_2)))]$$
$$+(1 - \delta) \log\{1 - \phi(\alpha'\tilde{Z})[1 - \exp(-\Lambda \exp(-(\beta_1 Z_1 + h(Z_2))))]\}. \quad (4.1)$$

With a slight abuse of notation, we write β_1 as β. Let the true value of h be h_0.

In our study, we assume h is smooth, more specifically a spline function. See assumption B1 below. Smoothness has also been assumed in Lam and Xue (2005). We propose using a penalty to control the smoothness of h, while a sieve approach was used in Lam and Xue (2005). An advantage of the penalized approach is that the degree of smoothness is controlled by a single number.

We consider the penalized maximum likelihood estimate (PMLE)

$$(\hat{\alpha}, \hat{\beta}, \hat{h}, \hat{\Lambda}) = argmax_{\alpha,\beta,h,\Lambda} \mathrm{E}_n l_2(X) - \lambda_n^2 J^2(h), \quad (4.2)$$

where λ_n is the data-dependent tuning parameter, $J^2(h) = \int_{Z_2}(h^{(s_0)}(Z_2))^2 dZ_2$ is the penalty on smoothness, and $h^{(s_0)}$ is the s_0^{th} derivative of h. In data analysis, one commonly takes $s_0 = 2$.

Remark 4. It can be shown that $\hat{h}$ defined in (4.2) is a spline function. Especially, suppose $\tilde{h}$ 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)).

4.1. Consistency and rate of convergence

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 $\|h - h_0\|_2^2 = \int_{Z_2}(h(Z_2) - h_0(Z_2))^2 dZ_2$. Beyond A1-A3, we also assume the following.

B1. (1) $h_0 \in \mathbb{H}_{s_0}$, the Sobolev space indexed by the order of derivative $s_0 \geq 1$;
(2) $\mathrm{P}(h_0(Z_2)) = 0$; (3) h_0 is uniformly bounded.
B2. For any $(\alpha, \beta, h, \Lambda)$ satisfying assumptions A1-A3 and B1,

$$\mathrm{E}[l_2(\alpha, \beta, h, \Lambda) - l_2(\alpha_0, \beta_0, h_0, \Lambda_0)] \leq -K_2 d^2((\alpha, \beta, h, \Lambda), (\alpha_0, \beta_0, h_0, \Lambda_0)),$$

with a fixed constant $K_2 > 0$.

An assumption similar to B1 is in Ma and Kosorok (2005a). Assumption B2 corresponds to A4. If for the tuning parameter λ_n we can assume

B3. $\lambda_n = O_p(n^{-1/3})$,

then the following result holds.

Lemma 7. (Consistency and rate of convergence). *Under* A1$-$A3 *and* B1$-$B3, $d((\hat{\alpha}, \hat{\beta}, \hat{h}, \hat{\Lambda}), (\alpha_0, \beta_0, h_0, \Lambda_0)) = O_p(n^{-1/3})$ and $J(\hat{h}) = O_p(1)$.

Lemma 7 shows that with presence of the nonparametric covariate effect, Λ_0 can still be estimated with the optimal convergence rate. 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 Cox model, h_0 can only be estimated at the $n^{1/3}$ rate. The overall entropy is driven by the entropy of Λ, resulting in an overall rate of $n^{1/3}$. We note that assumption B3 is different from the commonly assumed $\lambda_n = O_p(n^{-s_0/2s_0+1})$ (Wahba (1990)). Modifying this assumption cannot improve the convergence rate of $\hat{h}$, as can be seen from the proof.

4.2. Fisher information

Due to presence of the second nonparametric parameter h, standard information calculation based on orthogonal projection cannot be used. As an alternative, we apply the non-orthogonal projection (Sasieni (1992)). The score functions for α and β are

$$\dot{l}_{2\alpha} = \left[\frac{\delta}{\phi} - \frac{(1-\delta)g_2}{1 - \phi g_2}\right]\dot{\phi}\tilde{Z}, \quad \dot{l}_{2\beta} = \left[\frac{\delta}{g_2} - \frac{(1-\delta)\phi}{1 - \phi g_2}\right]g_3\Lambda_0 Z_1,$$

where $g_2(\alpha, \beta, h, \Lambda) = 1 - \exp[-\Lambda\exp(-(\beta Z_1 + h(Z_2)))]$, and $g_3(\alpha, \beta, h, \Lambda) = -\exp(-(\beta Z_1 + h(Z_2)))\exp[-\Lambda\exp(-(\beta Z_1 + h(Z_2)))]$. Let $\dot{l}_{2\alpha\beta} = (\dot{l}_{2\alpha}, \dot{l}_{2\beta})'$.

For $\eta \sim 0$, take $h_\eta = h + \eta\xi(Z_2)$ such that h_η still satisfies assumption B1. Denote the space generated by such ξ as $\mathbb{B}$. The score operator for h is $\dot{l}_{2h}(\xi) = [(\delta/g_2) - ((1-\delta)\phi)/(1 - \phi g_2)]\Lambda g_3 \xi$.

For the baseline Λ, consider $\Lambda_u = \Lambda + ua(c)$, with $u \sim 0$ and $a \in \mathbb{A}$. The score operator for Λ is $\dot{l}_{2\Lambda}(a) = -[(\delta/g_2) - ((1-\delta)\phi)/(1 - \phi g_2)]g_3 a = \tilde{l}_{2\Lambda} a$.
Step 1. We first project $\dot{l}_{2\alpha\beta}$ onto the space generated by the $\dot{l}_{2\Lambda}(a)$. We need to find a "direction" $a^* \in \mathbb{A}$ such that $\dot{l}_{2\alpha\beta} - \dot{l}_{2\Lambda}(a^*)\perp\dot{l}_{2\Lambda}(a)$ for all $a \in \mathbb{A}$. This is equivalent to requiring $\mathrm{E}[(\dot{l}_{2\alpha\beta} - \tilde{l}_{2\Lambda}a^*)\tilde{l}_{2\Lambda}a] = 0$. Using the standard projection approach, we can see that $a^* = [(\mathrm{E}(\dot{l}_{2\alpha\beta}\tilde{l}_{2\Lambda}|C))/(\mathrm{E}(\tilde{l}_{2\Lambda}\tilde{l}_{2\Lambda}|C))]$. Hence we have $\dot{l}_{2\alpha\beta} - \dot{l}_{2\Lambda}(a^*) = \dot{l}_{2\alpha\beta} - \tilde{l}_{2\Lambda}[(\mathrm{E}(\dot{l}_{2\alpha\beta}\tilde{l}_{2\Lambda}|C))/(\mathrm{E}(\tilde{l}_{2\Lambda}\tilde{l}_{2\Lambda}|C))]$.

Step 2. We now project $\dot{l}_{2h}(\xi)$ onto the space generated by $\dot{l}_{2\Lambda}(a)$, using calculations similar to those in Step 1. Denote the least favorable direction as $b^* \in \mathbb{A}$ and we have $\dot{l}_{2h}(\xi) - \dot{l}_{2\Lambda}(b^*) = \dot{l}_{2h}(\xi) - \tilde{l}_{2\Lambda}[(\mathrm{E}(\dot{l}_{2h}\tilde{l}_{2\Lambda}|C))/(\mathrm{E}(\tilde{l}_{2\Lambda}\tilde{l}_{2\Lambda}|C))]$.

Step 3. We project the space generated by $\dot{l}_{2\alpha\beta} - \dot{l}_{2\Lambda}(a^*)$ onto the space generated by $\dot{l}_{2h}(\xi) - \dot{l}_{2\Lambda}(b^*)$, which is equivalent to finding $\xi^* \in \mathbb{B}$ such that

$$
\mathrm{E}\Bigg\{ \left(\left[\dot{l}_{2\alpha\beta} - \tilde{l}_{2\Lambda}\frac{\mathrm{E}(\dot{l}_{2\alpha\beta}\tilde{l}_{2\Lambda}|C)}{\mathrm{E}(\tilde{l}_{2\Lambda}\tilde{l}_{2\Lambda}|C)} \right] - \left[\dot{l}_{2h}(\xi^*) - \tilde{l}_{2\Lambda}\frac{\mathrm{E}(\dot{l}_{2h}(\xi^*)\tilde{l}_{2\Lambda}|C)}{\mathrm{E}(\tilde{l}_{2\Lambda}\tilde{l}_{2\Lambda}|C)} \right] \right)
$$
$$
\times \left(\dot{l}_{2h}(\xi) - \tilde{l}_{2\Lambda}\frac{\mathrm{E}(\dot{l}_{2h}(\xi)\tilde{l}_{2\Lambda}|C)}{\mathrm{E}(\tilde{l}_{2\Lambda}\tilde{l}_{2\Lambda}|C)} \right) \Bigg\} = 0 \tag{4.3}
$$

for any $\xi \in \mathbb{B}$. Let $\dot{l}_2 = [\dot{l}_{2\alpha\beta} - \tilde{l}_{2\Lambda}((\mathrm{E}(\dot{l}_{2\alpha\beta}\tilde{l}_{2\Lambda}|C))/(\mathrm{E}(\tilde{l}_{2\Lambda}\tilde{l}_{2\Lambda}|C)))] - [\dot{l}_{2h}(\xi^*) - \tilde{l}_{2\Lambda}((\mathrm{E}(\dot{l}_{2h}(\xi^*)\tilde{l}_{2\Lambda}|C))/(\mathrm{E}(\tilde{l}_{2\Lambda}\tilde{l}_{2\Lambda}|C)))]$. We assume the following.

B4. There exists $\xi^* \in \mathbb{B}$ such that (4.3) is satisfied and $I_2 = \mathrm{E}(\dot{l}_2\dot{l}_2')$ is positive definite and component-wise bounded.

Lemma 8. (Fisher Information). *Under* A1−A3, B1−B2 *and* B4, I_2 *is the efficient information matrix for* (α, β).

4.3. Asymptotic normality and inference

Lemma 9. (Asymptotic normality and efficiency). *Under* A1−A3 *and* B1−B4, $\sqrt{n}(\hat{\alpha} - \alpha_0, \hat{\beta} - \beta_0) \rightarrow_d N(0, I_2^{-1})$.

Despite the slower than standard convergence rate of $\hat{h}$, the estimate of (α, β) is still asymptotically normal and efficient.

Remark 5. We note that the Fisher Information matrix I_2 for the partly linear Cox model does not have a closed form. A plug-in variance estimate does not seem feasible. However, as noted in Ma and Kosorok (2005b), the weighted bootstrap is still valid for penalized estimates with more than one nonparametric parameters.

5. Numerical Study

5.1. Computational algorithm

Under the partly linear model, the proposed PMLE requires maximization over multiple parametric parameters and two nonparametric parameters subject to constraints, which cannot be realized using any existing software. For estimating the spline function h, we take the sieve approach proposed in Xiang and Wahba (1997), that states that an estimate with the number of basis functions

growing at the rate of $n^{1/5}$ can achieve the same asymptotic precision as the full space. For data analysis, we suggest at least 20 basis functions. In our study, we choose equally spaced knots, and utilize B-spline basis functions. Once the basis functions are chosen, maximization over the nonparametric h becomes a parametric maximization problem. Maximization over α, β and the regression coefficients in h can be achieved via the Newton-Raphson method, or using functions like *optim* in R.

Maximization over the non-decreasing baseline function Λ is achieved with the pool-adjacent-violator (PAV) approach, investigated in Barlow, Bartholomew, Bremner and Brunk (1972). Application of the PAV in current status data study can also be found in van der Laan and Jewell (2001).

Simultaneous maximization over all parameters is difficult. So we consider the following iterative procedure.

1. Initialize $\hat{\alpha}, \hat{\beta}$ and $\hat{h}$ at zero.
2. With the current estimate of (α, β, h), maximize over Λ using the PAV approach.
3. With the current estimate of Λ, maximize over (α, β, h) using the *optim* function in R.
4. Repeat Steps 2 and 3 until convergence.

Research code written in R is available upon request. Our limited numerical studies show that convergence is usually achieved in 20 iterations. For a dataset with sample size 400, one estimation takes less than 2 minutes.

The proposed PMLE involves the tuning parameter λ_n. Asymptotically, we assume λ_n satisfies B3. For data analysis, we propose setting $\lambda_n = \tau \times n^{-1/3}$, where $\tau = 0.1, 0.2, \ldots, 10.0$. We propose using five-fold cross validation (Wahba (1990)) and searching over τ to determine its optimal value (and hence the optimal λ_n). Our limited numerical study shows that optimal tunings can usually be found in the searching interval.

5.2. Simulation study

We conduct simulation studies to assess the finite sample performance of the proposed estimates. For each model, we consider sample size $n = 200$ and 400. We assume ϕ is the logistic regression function, and take the cumulative baseline function $\Lambda_0(T) = T$. Inference is based on the proposed weighted bootstrap with exponential weights and $B = 500$. We consider the following four models.

Model 1. $\alpha_0 = (0, 1, 1)$ and $\beta_0 = (1, -1)$; $Z_1 = 0$ or 2 with probabilities $1/2$; $Z_2 \sim Unif[0, 2]$. The censoring time is truncated $\exp(0.5)$ distributed with

Table 1. Simulation study: summary statistics based on 500 replicates. *sd*: standard deviations of $\hat{\alpha}$ and $\hat{\beta}$; $\hat{sd}$: standard deviations of $\hat{\alpha}^* - \hat{\alpha}$ and $\hat{\beta}^* - \hat{\beta}$.

		α_1	α_2	β_1	β_2
		Model 1			
$n = 200$	bias	0.006	0.054	0.021	0.049
	sd	0.056	0.177	0.128	0.181
	$\hat{sd}$	0.055	0.186	0.110	0.165
$n = 400$	bias	0.007	0.036	0.004	-0.025
	sd	0.041	0.125	0.091	0.130
	$\hat{sd}$	0.035	0.129	0.088	0.129
		Model 2			
$n = 200$	bias	0.034	0.037	0.009	-0.067
	sd	0.121	0.185	0.155	0.157
	$\hat{sd}$	0.114	0.187	0.158	0.154
$n = 400$	bias	0.017	0.031	0.014	-0.041
	sd	0.085	0.133	0.111	0.106
	$\hat{sd}$	0.087	0.122	0.113	0.101
		Model 3			
$n = 200$	bias	0.033	0.023	-0.070	–
	sd	0.131	0.175	0.127	–
	$\hat{sd}$	0.118	0.165	0.112	–
$n = 400$	bias	0.022	0.012	0.032	–
	sd	0.093	0.113	0.092	–
	$\hat{sd}$	0.104	0.123	0.088	–
		Model 4			
$n = 200$	bias	0.048	-0.003	-0.044	–
	sd	0.154	0.153	0.131	–
	$\hat{sd}$	0.144	0.162	0.119	–
$n = 400$	bias	0.024	0.016	-0.031	–
	sd	0.113	0.103	0.099	–
	$\hat{sd}$	0.108	0.111	0.091	–

an upper bound of 7. The probability of cure is 0.17; for subjects not cured, the censoring rate is 0.40.

Model 2. $\alpha_0 = (0, 1, 1)$ and $\beta_0 = (1, -1)$; $Z_1 \sim Unif[0, 2]$ and $Z_2 \sim Unif[0, 2]$. The censoring distribution is the same as in Model 1. The probability of cure is 0.15; for subjects not cured, the censoring rate is 0.35.

Model 3. $\alpha_0 = (0, 2, 1)$ and $\beta_0 = -1$; $Z_1 = 0$ or 2 with probabilities $1/2$; $Z_2 \sim Unif[-1, 3]$; $h_0(Z_2) = \sin(\pi Z_2)$. The censoring time is truncated $\exp(1)$ distributed with an upper bound of 4.5. The probability of cure is 0.17; for subjects not cured, the censoring rate is 0.30.

Model 4. $\alpha_0 = (0, 2, 1)$ and $\beta_0 = -1$; $Z_1 \sim Unif[0, 2]$; $Z_2 \sim Unif[-1, 3]$; $h_0(Z_2) = \sin(\pi Z_2)$. The censoring is the same as in Model 3. The probability of cure is 0.10; for subjects not cured, the censoring rate is 0.30.
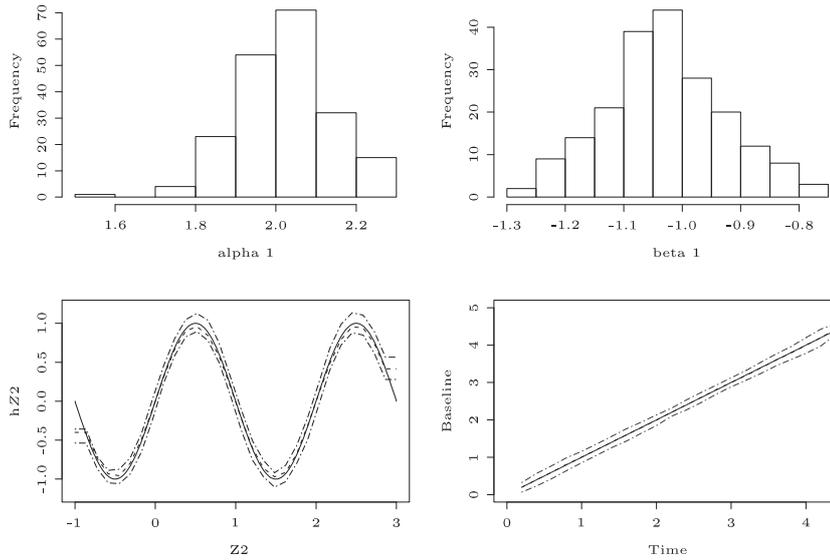
Figure 5.1. Simulation study (Model 4, sample size 400). Upper left: histogram of the second component of $\hat{\alpha}$; Upper right: histogram of $\hat{\beta}$; Lower left: estimate of h_0, mean estimate and point-wise 95% confidence interval; Lower right: estimate of Λ_0, mean estimate and point-wise 95% confidence interval. The solid lines are the true values.

Models 1 and 2 are linear Cox models, while Models 3 and 4 have non-parametric covariate effects. 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 there that the estimates have small biases. The standard deviations of the estimates shrink as the sample size increase from 400 to 800 by approximately $\sqrt{2}$. The weighted bootstrap estimates of standard errors are close to the estimates' standard deviations.

In Figure 5.1, we show the simulations plots for Model 4 with sample size 400. The top panels show the histograms of estimated α (the second component) and β. We can see that the estimates have distributions close to normal. In the bottom panels, we show the point-wise means of estimated h and Λ; we also show the point-wise 95% confidence intervals. The point-wise confidence intervals are informal and provided simply to show the variations of the estimates. We can see that the mean $\hat{h}$ matches the true h_0 very well, with small variation. The estimated h is less satisfactory when it is close to boundaries and there are fewer data points.

5.3. Calcification study

In the Calcification study, patients were examined at times ranging from 0 to 36 months after the hydrogel intraocular lenses implantation. The severity of calcification was graded on a discrete scale ranging from 0 to 4, with severity ≤ 1 classified as "not calcified". The clinical factors of scientific interest include gender, incision length, and age at implantation. The dataset contains 379 records. We exclude the one record with a missing measurement. Of $n = 378$ subjects, 48 experienced calcification.

Let $Z_1 = incision\ length$, $Z_2 = gender\ (male)$, $Z_3 = age/10$, and $Z = (Z_1, Z_2, Z_3)$. We assume the logistic model

$$prob(not\ cure) = \frac{\exp(\alpha_1 + \alpha_2 Z_3)}{1 + \exp(\alpha_1 + \alpha_2 Z_3)}. \tag{5.1}$$

For subjects susceptible to calcification, the conditional cumulative hazard satisfies the linear Cox model

$$\Lambda(T|Z) = \Lambda(T)\exp(-(\beta_1 Z_1 + \beta_2 Z_2 + \beta_3 Z_3)), \tag{5.2}$$

or the partly linear Cox model

$$\Lambda(T|Z) = \Lambda(T)\exp(-(\beta_1 Z_1 + \beta_2 Z_2 + h(Z_3))). \tag{5.3}$$

The proposed MLE and PMLE are used for estimation. Inference is based on the weighted bootstrap with 500 realizations of random exponential weights.

Under (5.1) and (5.2), the MLEs are $\hat{\alpha}_1 = -0.998(0.078)$, $\hat{\alpha}_2 = 0.074(0.077)$, $\hat{\beta}_1 = -0.352(0.126)$, $\hat{\beta}_2 = 0.230(0.114)$, and $\hat{\beta}_3 = 0.211(0.075)$, where the values in parentheses are the corresponding bootstrap standard deviation estimates. We can see that older people are less likely to be cured, although for those who are not cured, older people have a smaller risk of calcification. Such results appear contradictive. However, we note that the age effect in the cure model is not significant, so its effect on the cure rate is not conclusive. All three covariates being considered have significant effects on the survival risk for susceptible subjects: increase in incision length, being female, and decrease in age lead to an increase in calcification hazard.

We also consider models (5.1) and (5.3). With the five-fold cross validation, $\tau = 0.5$ (hence $\lambda_n = 0.5n^{1/3}$) is selected. The PMLE estimates are $\hat{\alpha}_1 = -1.004(0.079)$, $\hat{\alpha}_2 = -0.040(0.032)$, $\hat{\beta}_1 = -0.228(0.111)$, and $\hat{\beta}_2 = 0.399(0.094)$. We can see that the estimates are considerably different from their counterparts under the linear Cox model. This is partly caused by the fact that the age effect is significantly different from linear (Figure 5.2). We also show the estimated
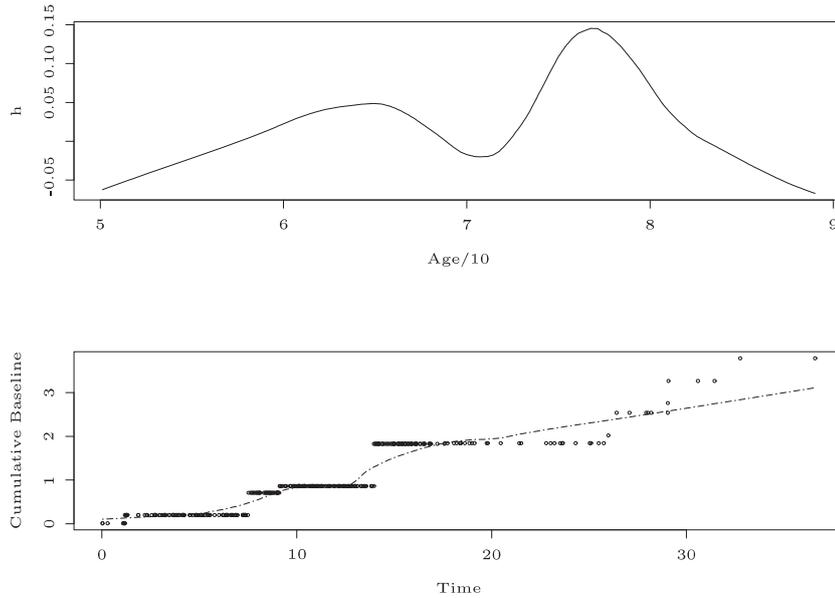
Figure 5.2. Calcification data under the partly linear model. Estimation of h_0 and Λ_0 (with lowess smoother).

cumulative baseline (with its lowess smoother) in Figure 5.2. Under (5.1) and (5.3), we see that older people are more likely to be cured, although this effect is not significant. For susceptible subjects, increase in incision length and being female lead to an increase in calcification hazard. The effect of age on calcification risk is nonlinear – it first increases with age and then decreases, with a local minimum at age 71.

6. Remarks

Our focus has been on linear and partly linear models. A more flexible model has $f(Z) = h_1(Z_1) + h_2(Z_2)$. If we make the assumption that h_1 is also a spline function, then we can consider a doubly penalized estimate, with penalties on the smoothness of both h_1 and h_2. Asymptotically, as long as $S(u_C) < 1$, the proposed approach holds. However we note that for finite sample sizes, testing the existence of the cured subgroup may be essential (Maller and Zhou (1992)). We postpone this discussion to a future study. For the Calcification data analysis, we only consider two specific Cox models and one link function. These models are motivated by Lam and Xue (2005). For any practical data analysis, robustness of the estimates and performance under model misspecification are of interest, but beyond the scope of the present paper.

Acknowledgement

References

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

Barlow, R. E., Bartholomew, D. J., Bremner, J. M. and Brunk, H. D. (1972). *Statistical Inference Under Order Restrictions.* Wiley, New York.

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

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

Cox, D. R. (1972). Regression models and life tables (with discussion). *J. Roy. Statist. Soc. Ser. B* **34**, 187-22.

Fang, H. B, Li, G. and Sun, J. (2005). Maximum likelihood estimation in a semiparametric logistic/proportional-hazards mixture model. *Scand. J. Statist.* **32**, 59-75.

Farewell, V. T. (1986). Mixture models in survival analysis: are they worth the risk? *Canad. J. Statist.* **14**, 257-262.

Gart, J. J., Krewski, D., Lee, P. N., Tarone, R. E. and Wahrendorf, J. (1986). *Statistical Methods in Cancer Research, Volume III, The Design and Analysis of Longterm Animal Experiments.* IARC Scientific Publications N0. 79. International Agency for Research on Cancer, Lyon.

Groeneboom, P. and Wellner, J. A. (1992). *Information Bounds and Nonparametric Maximum Likelihood Estimation.* Birkhauser, Basel.

Grummer-Strawn, L. M. (1993). Regression analysis of current status data: An application to breast feeding. *J. Amer. Statist. Assoc.* **88**, 758-765.

Huang, J. (1996). Efficient estimation for the proportional hazard model with interval censoring. *Ann. Statist.* **24**, 540-568.

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

Kuk, A. Y. C. and Chen, C. H. (1992). A mixture model combining logistic regression with proportional hazards regression. *Biometrika* **79**, 531-541.

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

Li, C. S., Taylor, J. M. G. and Sy, J. P. (2001). Identifiability of cure models. *Statist. Probab. Lett.* **54**, 389-395.

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

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

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

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

Maller, R. A. and Zhou, S. (1992). Estimating the proportion of immunes in a censored sample. *Biometrika* **79**, 731-739.

Peng, Y. and 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. *Scand. J. Statist.* **19**, 215-233.

Sy, J. P. and Taylor, J. M. G. (2000). Estimation in a Cox proportional hazards cure model. *Biometrics* **56**, 227-236.

Sy, J. P. and Taylor, J. M. G. (2001). Standard errors for the Cox proportional hazards cure model. *Mathematical and Computer Modeling* **33**, 1237-1251.

Taylor, J. M. G. (1995). Semiparametric estimation in failure time mixture models. *Biometrics* **51**, 899-907.

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

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

van der Laan, M. and Jewell, N. P. (2001). The NPMLE for doubly censored current status data. *Scand. J. Statist.* **28**, 537-547.

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

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics.* Springer, New York.

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

Xiang, D. and Wahba, G. (1997). Approximate smoothing spline methods for large data set in the binary case. *Proceedings of the 1997 ASA Joint Statistical Meetings*, Biometrics Section, 94-98.

Xue, H., Lam, K. F. and Li, G. (2004). Sieve maximum likelihood estimator for semiparametric regression models with current status data. *J. Amer. Statist. Assoc.* **99**, 346-356.

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

Department of Epidemiology and Public Health, Yale University, New Haven, CT 06520, U.S.A.

E-mail: shuangge.ma@yale.edu