Statistica Sinica **20** (2010), 1165-1181

MIXED CASE INTERVAL CENSORED DATA WITH A CURED SUBGROUP

Shuangge Ma

Yale University

Abstract: Mixed case interval censored data arise when the event time of interest is only known to lie in an interval obtained from a sequence of k random examinations, where k is a random integer. In this article, we consider mixed case interval censored data with a cured subgroup, where subjects in this subgroup are not susceptible to the event of interest. Such data may be encountered in medical and demographical studies with longitudinal followup, where the population of interest is composed of heterogeneous subjects. We propose using a semiparametric two-part model, where the first part is a generalized linear model that describes the probability of cure, and the second part is a Cox model that describes the event time for susceptible subjects. We study maximum likelihood estimation of this two-part model. Finite sample properties, an effective computational algorithm, and inference with the weighted bootstrap are investigated. Asymptotic properties, including identifiability, consistency, and weak convergence, are established. We conduct simulations and analyze the HDSD study using the proposed approach.

Key words and phrases: Cure rate, interval censoring, semiparametric model.

1. Introduction

We have been partly motivated by the analysis of an experiment conducted by NASA on decompression sickness. The working dataset is extracted from the NASA's Hypobaric Decompression Sickness Data Bank (Conkin, Bedahl and Van Liew (1992)) and is be referred to as the HDSD hereafter. The presence of gas bubbles in venous blood is associated with an increased risk of decompression sickness (DCS) in hypobaric environments. A high grade of venous gas emboli (VGE) can be a precursor to serious DCS. Therefore, it is important to model the time to onset of grade IV VGE in order to predict the situations in which it is most likely to occur. The dataset has records from volunteer subjects undergoing denitrogenation test procedures prior to being exposed to a hypobaric environment. Each test involved one decompression, where the subject pre-breathed 100% oxygen at site pressure prior to exposure in the altitude chamber. For each subject, the time to onset of grade IV VGE and values of several covariates were measured. The onset time, if it occurred, was recorded only as being contained within a time interval. When the experiment was conducted, for a subject there

might be multiple examination times. That is, the onset time is case k interval censored, where k may vary across subjects. Such type of data has been referred to as "mixed case" interval censored data in the literature (Schick and Yu (2000); Song (2004); Sen and Banerjee (2007)). Beyond the complexity caused by interval censoring, it has been suggested that "some individuals would never get grade IV VGE no matter how long they remained in the hypobaric chamber" (Thompson and Chhikara (2003)). Plot of the NPMLE of the survival function clearly shows a plateau (plot available upon request). A cure rate model that allows a subgroup of subjects to be immune from the event of interest is thus warranted.

Interval censored data arise naturally in medical, biological, and demographical studies. With interval censored data the event time of interest cannot be directly observed, it is only known to lie in an interval obtained from a sequence of examinations (censoring). One way to characterize interval censored data is by using the number of random censoring times (denoted as k). Case I interval censored data ($k = 1$), also known as current status data, has been investigated in Huang (1996), Lin, Oakes and Ying (1998), Xue, Lam and Li (2004), Ma and Kosorok (2005a), and others. Nonparametric modeling of case II interval censored data ($k = 2$) has been studied in Groeneboom and Wellner (1992). With $k > 1$, it has been pointed out that case k interval censoring is not realistic. For example, when it is known that the event has occurred by examination $k'(< k)$, there is no need to conduct further examinations. A more realistic model is mixed case interval censoring, where each subject is case k interval censored. Here, k is a random integer (as opposed to a fixed number). Mixed case interval censored data has been studied in Schick and Yu (2000), Song (2004), Sen and Banerjee (2007), and references therein.

Beyond its interval censoring nature, the HDSD data is difficult to analyze because of the subgroup not susceptible to the event. Such a phenomenon has been referred to as "cured" in statistical literature. For right censored data, studies of cure rate models include Kuk and Chen (1992), Lu and Ying (2004), Li, Taylor and Sy (2001), Taylor (1995), and others. On interval censored data with a cured subgroup, published studies include the semiparametric AFT model in Lam and Xue (2005), the Cox model in Ma (2009), the parametric models with a Bayesian estimator in Thompson and Chhikara (2003), the frailty models with a Bayesian estimator in Banerjee and Carlin (2004), and others.

In this article, we study semiparametric two-part models for mixed case interval censored data, where a cured subgroup is not susceptible to the event of interest. Such a data structure has been studied in Thompson and Chhikara (2003) using parametric models. Semiparametric models considered in this article can be much more flexible and, potentially, provide better descriptions of data.

The two-part model investigated here is similar to that in Ma (2009). However, the censoring schemes in the two articles are fundamentally different. Early studies have shown that the extension from case I to mixed case interval censored data is highly nontrivial. Compared with mixed case censoring studies in Schick and Yu (2000) and others, this study has been motivated by the analysis of HDSD data, which is heterogeneous and needs to be described using a mixture model. Since the present data structure may be frequently encountered and existing methodologies are not sufficient, the proposed study seems warranted.

Although we have been motivated by the HDSD data, applications of the proposed methodology go far beyond it. One family of studies that can be analyzed using the proposed method is the study of cancer recurrence after surgery. Time of cancer recurrence is usually not directly observable. Longitudinal monitoring is needed and can lead to interval censored data. In addition, susceptibility to cancer differs significantly among patients. Cure rate models are thus needed. We focus on methodological development in this article and defer exploration of its broader applications to future studies.

The rest of the article is organized as follows. Data and model setup are introduced in Section 2. The maximum likelihood estimate (MLE) is proposed in Section 3. Finite sample properties, an effective computational algorithm, and inference are investigated. Asymptotic properties are established in Section 4. Simulation studies in Section 5 demonstrate satisfactory finite sample performance of the proposed approach. Analysis of the HDSD data is presented in Section 6. We conclude with discussion in Section 7.

2. Data and Model

Let T be the event time of interest. Under mixed case interval censoring, the censoring is determined via a two-step procedure. In the first step, k, the number of censoring times, is determined. In the second step, the observation is determined by a case k interval censoring model. Let $(\tilde{U}_1, \ldots, \tilde{U}_k)$ be the random censoring times. As pointed out in previous studies, only (U, V) is relevant to statistical modeling, where (U, V) is the shortest interval such that $U < T \leq V$. Of note, it is possible that $U = 0$ or $V = \infty$. For convenience of notation, we introduce the left and interval censoring indicators, defined on values of U and V as $\delta_1 = I(U = 0)$ and $\delta_2 = I(U > 0 \cap V < \infty)$, where I is the indicator function. Here we assume $P(\tilde{U}_1 = 0) = 0$, i.e., the first examination happens after starting of the study; this is reasonable for most biomedical studies. To account for the possibility of cure, we introduce the *unobservable* cure indicator Y: $Y = 1$ if the subject is cured or immune ($T = \infty$) and $Y = 0$ otherwise. Let Z be the length d_1 vector of covariates that are associated with the cure probability, and X be the length d_2 covariates that are associated with the survival for susceptible

subjects. X and Z may have no, partial, or full overlap. Observation for a single subject consists of $D = (U, V, Z, X)$.

We model the data structure described above with a two-part model, where the first part models the cure probability and the second models the survival time for susceptible subjects. More specifically, we model the cure probability using a generalized linear model with a known link function. The most widely used link function is the logit link, where

$$p(Z) = P(Y = 1|Z) = \frac{\exp(\alpha + \beta' Z)}{1 + \exp(\alpha + \beta' Z)}. \tag{2.1}$$

Here α is the unknown intercept, β is the unknown length d_1 regression coefficient, and β' is the transpose of β.

In the second part of the two-part model, we assume the Cox model for subjects susceptible to the event of interest. Under the Cox model, the conditional hazard function is $\lambda(t|X) = \lambda(t)\exp(\theta' X)$, where $\lambda(t)$ is the baseline hazard function, and θ is the length d_2 regression coefficient. In terms of the cumulative hazard function,

$$\Lambda(t|X) = \int_0^t \lambda(s|X)ds = \Lambda(t)\exp(\theta' X). \tag{2.2}$$

Under models (2.1) and (2.2), the conditional survival function is

$$S(t|Z, X) = p(Z) + (1 - p(Z))\exp(-\Lambda(t|X)). \tag{2.3}$$

$S(\infty|Z, X) = p(Z)$ and is thus improper if $p(Z) > 0$.

Two-part models have been extensively used in analysis of heterogeneous data. Compared with the latent cause model in Thompson and Chhikara (2003), two-part models make no assumption on the latent cause, and can be preferred for data such as the HDSD.

3. Maximum Likelihood Estimation

Consider the model defined in (2.3). Under mixed case interval censoring, the log-likelihood function for a single observation is

$$\begin{aligned} l(\alpha, \beta, \theta, \Lambda) &= \delta_1 \log[(1 - p(Z))(1 - \exp(-\Lambda(V)\exp(\theta' X)))] \\ &\quad + \delta_2 \log[(1 - p(Z))(\exp(-\Lambda(U)\exp(\theta' X)) - \exp(-\Lambda(V)\exp(\theta' X)))] \\ &\quad + (1 - \delta_1 - \delta_2)\log[p(Z) + (1 - p(Z))\exp(-\Lambda(U)\exp(\theta' X)))]. \end{aligned} \tag{3.1}$$

Assume there are n i.i.d. copies of D. We consider the MLE

$$(\hat{\alpha}, \hat{\beta}, \hat{\theta}, \hat{\Lambda}) = argmax \mathrm{E}_n l(\alpha, \beta, \theta, \Lambda), \tag{3.2}$$

where E_n is the empirical measure.

3.1. Assumptions

A1. k and $(\tilde{U}_1, \ldots, \tilde{U}_k)$ are independent of (X, Z).

A2. (a) The distribution of Z is not concentrated on any proper subspace of $\mathbb{R}^{d_1}$; Z belongs to a bounded subset of $\mathbb{R}^{d_1}$. (b) The distribution of X is not concentrated on any proper subspace of $\mathbb{R}^{d_2}$; X belongs to a bounded subset of $\mathbb{R}^{d_2}$.

A3. (a) There exists a positive η such that $P(V - U \geq \eta) = 1$. (b) The union of the support for $V|\delta_1 = 1$, $U|\delta_2 = 1$, $V|\delta_2 = 1$, and $U|\delta_1 = \delta_2 = 0$ is an interval $[\tau_0, \tau_1]$ with $0 < \tau_0 < \tau_1 < \infty$.

A4. (a) (α, β, θ) belongs to a compact subset of $\mathbb{R}^{1+d_1+d_2}$. (b) There exists $M > 0$, such that $1/M < \Lambda(\tau_0) < \Lambda(\tau_1) < M$.

The independence assumption A1 has been commonly made in interval censoring studies; A2 is needed in the identifiability and consistency proofs; A3 has it that the examination times are bounded away from each other, which rules out accurately observed event times. As pointed out by Kim (2003), if a proportion of the event times can be accurately observed, properties of the MLE can be fundamentally different; we focus on the scenario with no accurately observed event times that better describes the HDSD and most biomedical studies. In addition, we only consider bounded examination times. Such an assumption is reasonable, considering that most (if not all) biomedical studies are conducted within finite time periods. A byproduct of A3 is that k is bounded. Such an assumption is stronger than the assumption $\mathrm{E}(k) < \infty$ in Schick and Yu (2000), although in practice such a difference has a negligible impact. The boundedness assumption (A4) is also commonly made with interval censored data, and is used in the consistency and weak convergence proofs.

3.2. Finite sample properties

Under A2 and A4, the MLE defined in (3.2) exists. For uniqueness, we further specify that $\hat{\Lambda}$ is right continuous, piecewise constant, and with possible discontinuities only at $\{C_j : \ j = 1 \ldots m\}$, which are unique values of $\{U_i, V_i : i = 1 \ldots n\}$.

First, we note that the MLE satisfies

$$\left.\frac{\partial \mathrm{E}_n l}{\partial \alpha}\right|_{\hat{\alpha},\hat{\beta},\hat{\theta},\hat{\Lambda}} = \left.\frac{\partial \mathrm{E}_n l}{\partial \beta}\right|_{\hat{\alpha},\hat{\beta},\hat{\theta},\hat{\Lambda}} = \left.\frac{\partial \mathrm{E}_n l}{\partial \theta}\right|_{\hat{\alpha},\hat{\beta},\hat{\theta},\hat{\Lambda}} = 0. \tag{3.3}$$

This follows from the definition of MLE, the differentiability of likelihood function, and the compactness assumptions. In (3.3), when we take the partial derivative with respect to one parameter, the other parameters are kept fixed and the "dependence" among estimates is ignored.

To investigate the properties of $\hat{\Lambda}$, we take $g_U = \exp(-\Lambda(U)\exp(\theta' X))$, $g_V = \exp(-\Lambda(V)\exp(\theta' X))$, and define the processes

$$\begin{aligned} W_\Lambda(t) = &\sum_{i=1}^{n} \frac{\delta_{1i}\exp(\theta' X_i) g_V(D_i)}{1-g_V(D_i)} I(V_i \le t) + \frac{\delta_{2i}\exp(\theta' X_i) g_V(D_i)}{g_U(D_i)-g_V(D_i)} I(V_i \le t) \\ &- \frac{\delta_{2i}\exp(\theta' X_i) g_U(D_i)}{g_U(D_i)-g_V(D_i)} I(U_i \le t) \\ &- \frac{(1-\delta_{1i}-\delta_{2i})(1-p(Z_i))\exp(\theta' X_i) g_U(D_i)}{p(Z_i)+(1-p(Z_i))g_U(D_i)} I(U_i \le t), \\ G_\Lambda(t) = &\, W_\Lambda^2(t), \\ Q_\Lambda(t) = &\, W_\Lambda(t) + \int \Lambda dG_\Lambda. \end{aligned}$$

Let $\{C_{(j)}\}$ be the ordered $\{C_j\}$. For $(\alpha, \beta, \theta) = (\tilde{\alpha}, \tilde{\beta}, \tilde{\theta})$, let $\tilde{\Lambda}$ be the left derivative of the greatest convex minorant of the self-induced cumulative sum diagram formed by the points $(0,0)$ and $(G_{\tilde{\Lambda}}(C_{(j)}), Q_{\tilde{\Lambda}}(C_{(j)}))$. Then $\tilde{\Lambda}$ maximizes $\mathrm{P}_n l(\tilde{\alpha}, \tilde{\beta}, \tilde{\theta}, \Lambda)$ (as a function of Λ). The proof of this result is from Groeneboom and Wellner (1992), part II, Chapter 1.

3.3. Computational algorithm

The MLE defined in (3.2) does not have a closed, analytic form. We propose maximization using the following iterative algorithm that has been motivated by the finite sample properties presented in Section 3.2.

1. Initialize $(\hat{\alpha}, \hat{\beta}, \hat{\theta}) = (0 \dots 0)$.
2. With the current estimate $(\hat{\alpha}, \hat{\beta}, \hat{\theta})$, compute $\hat{\Lambda}$ by maximizing $\mathrm{E}_n l(\hat{\alpha}, \hat{\beta}, \hat{\theta}, \Lambda)$ as a function of Λ. This step of maximization can be achieved as follows.
   (a) With the current estimate $\hat{\Lambda}$, compute the left derivative of the greatest convex minorant of the cumulative sum diagram composed of $(0,0)$ and $(G_{\hat{\Lambda}}(C_{(j)}), Q_{\hat{\Lambda}}(C_{(j)}))$. This computation can be realized using available functions such as the *gcmlcm* in R. We refer to Robertson, Wright and Dykstra (1988) for detailed descriptions of computing the greatest convex minorant.

(b) Update $\hat{\Lambda}$ with the left derivative computed in (a). We note that the greatest convex minorant is a piece-wise linear function. Thus, we only need to compute the derivatives at a small number of points, and set $\hat{\Lambda}$ as right-continuous and piece-wise constant.

(c) Repeat Steps (a) and (b) until convergence.

3. With the estimated $\hat{\Lambda}$, maximize $\mathrm{E}_n l(\alpha, \beta, \theta, \hat{\Lambda})$ with respect to (α, β, θ). This can be achieved using the Newton-Raphson method, or built-in optimization functions such as the *optim* in R.

4. Repeat Steps 2 and 3 until convergence.

The empirical measurement of the log-likelihood function increases at each iteration. Under the compactness assumptions, the above algorithm always converges. Our numerical studies show that convergence can usually be achieved within 20 iterations.

3.4. Inference

With the proposed semiparametric model, inference of the parametric portion is of more interest than that of the nonparametric portion. Asymptotic studies in Section 4 show that $(\hat{\alpha}, \hat{\beta}, \hat{\theta})$ is $\sqrt{n}$ consistent and asymptotically normally distributed. However, the asymptotic variance calculation in Section 4.4 suggests that a plug-in estimate can be difficult. As an alternative we consider a weighted bootstrap, described as follows.

1. Generate $w_1 \dots w_n$, n i.i.d. positive random weights from a known distribution with $\mathrm{E}(W) = var(W) = 1$. (In the numerical studies, we used exp(1) distributed weights.)

2. Compute the weighted MLE $(\hat{\alpha}^*, \hat{\beta}^*, \hat{\theta}^*, \hat{\Lambda}^*) = argmax \sum_i w_i l(D_i)$.

3. Repeat Steps 1 and 2 B (e.g. 500) times.

The sample variance of $(\hat{\alpha}^*, \hat{\beta}^*, \hat{\theta}^*)$ can be used to estimate the variance of $(\hat{\alpha}, \hat{\beta}, \hat{\theta})$. In Step 2, the weighted MLE can be computed using an algorithm similar to the one described in Section 3.3 (by changing simple summations to weighted summations and keeping everything else the same).

The weighted bootstrap has been proposed as a generic inference tool for M-estimates with semiparametric models in Ma and Kosorok (2005b), and for U-estimates in Jin, Ying and Wei (2001). Of note, the weighted bootstrap methods for M-estimates and U-estimates share similar spirits and generate perturbations of objective functions by assigning random weights. However, the specific forms

of perturbations and assumptions on weights are different. With the weighted bootstrap, the weighted MLE needs to be computed many times. However, since only simple calculations are involved, the weighted bootstrap is computationally affordable.

4. Asymptotic Properties

4.1. Identifiability

We first establish identifiability of the model. Li, Taylor and Sy (2001) shows that with cure rate models, identifiability can be highly nontrivial. Here we make specific semiparametric model assumptions that, with the compactness assumptions, lead to identifiability of the model.

Lemma 1. *Under* A1−A4*, the proposed model is identifiable.*

Proof. Let μ be the probability measure induced by the joint distribution of U and V constrained on the interval $[\tau_0, \tau_1]$. Let $f(D;\alpha,\beta,\theta,\Lambda)$ be the probability density function of $D = (U, V, Z, X)$ measured at parameter value $(\alpha,\beta,\theta,\Lambda)$. The proposed model is identifiable if

$$\int \left(\sqrt{f(D;\alpha,\beta,\theta,\Lambda)} - \sqrt{f(D;\alpha^*,\beta^*,\theta^*,\Lambda^*)}\right)^2 d\mu = 0 \tag{4.1}$$

implies $(\alpha,\beta,\theta,\Lambda) = (\alpha^*,\beta^*,\theta^*,\Lambda^*)$.

Equation (4.1) leads to

$$\begin{aligned}&\left(1 - \frac{\exp(\alpha+\beta'Z)}{1+\exp(\alpha+\beta'Z)}\right)(1-\exp(-\Lambda(U)\exp(\theta'X)))\\ &\quad = \left(1 - \frac{\exp(\alpha^*+\beta^{*\prime}Z)}{1+\exp(\alpha^*+\beta^{*\prime}Z)}\right)(1-\exp(-\Lambda^*(U)\exp(\theta^{*\prime}X))),\end{aligned}$$

which implies that

$$1-\exp(-\Lambda(U)\exp(\theta'X)) = c(Z)(1-\exp(-\Lambda^*(U)\exp(\theta^{*\prime}X))),$$

where c is a function of Z only. Take partial derivatives of both sides of the last equality with respect to U:

$$\lambda(U)\exp(-\Lambda(U)\exp(\theta'X)) = c(Z)\lambda^*(U)\exp(-\Lambda^*(U)\exp(\theta^{*\prime}X)).$$

Take logarithm and then take the partial derivative with respect to X:

$$-\Lambda(U)\exp(\theta'X)\theta = \frac{1}{c(Z)}\frac{\partial c(Z)}{\partial X} - \Lambda^*(U)\exp(\theta^{*\prime}X)\theta^*.$$

Take the partial derivative with respect to U again to get

$$\frac{\Lambda(U)}{\Lambda^*(U)} = \frac{\exp(\theta' X)\theta' X}{\exp(\theta^{*\prime} X)\theta^{*\prime} X}.$$

Under A2, this last implies that $\Lambda = \Lambda^*$ and $\theta = \theta^*$. In a similar manner, it can be proved that $\alpha = \alpha^*$ and $\beta = \beta^*$.

4.2. Consistency

Let $m_\Lambda(\Lambda_1, \Lambda_2) = \int[(\Lambda_1(u) - \Lambda_2(u))^2 + (\Lambda_1(v) - \Lambda_2(v))^2]^{1/2} d\mu$ and denote the unknown true parameter value by $(\alpha_0, \beta_0, \theta_0, \Lambda_0)$. The consistency result can be summarized as follows.

Lemma 2. *Under* A1−A4, $(\hat{\alpha}, \hat{\beta}, \hat{\theta}) \to_{a.s.} (\alpha_0, \beta_0, \theta_0)$ *and* $m_\Lambda(\hat{\Lambda}, \Lambda_0) = o_P(1)$.

Proof. From the definition of MLE,

$$\mathrm{E}_n l(\hat{\alpha}, \hat{\beta}, \hat{\theta}, \hat{\Lambda}) \geq \mathrm{E}_n l(\alpha_0, \beta_0, \theta_0, \Lambda_0). \tag{4.2}$$

Under the compactness assumptions, the right side of (4.2) is bounded below. If $(\hat{\alpha}, \hat{\beta}, \hat{\theta})$ is not bounded, then $\mathrm{E}_n l(\hat{\alpha}, \hat{\beta}, \hat{\theta}, \hat{\Lambda}) \to -\infty$. Consider for example if $\hat{\alpha} \to +\infty$, then $p(Z) \to 1$. In the log-likelihood function, the first two terms $\to -\infty$, whereas the third term $\to 0$. We can repeat this argument and conclude boundedness of $(\hat{\alpha}, \hat{\beta}, \hat{\theta})$. The functional set $\{\exp(-\Lambda)\}$ is monotone and bounded. It is thus compact with respect to the vague topology. Consistency then follows from Theorem 5.14 of van der Vaart (1998).

Results presented in Lemmas 2 and 3 establish consistency of $\hat{\Lambda}$ in the L_2 sense. Consistency under other norms may require different, possibly stronger, assumptions (Schick and Yu (2000)). Since it is not the focus of this study, we do not pursue consistency under other norms. On a special note, the L_2 consistency does not lead to uniform consistency. Specifically, we expect that $\hat{\Lambda}$ is not consistent at τ_0 and τ_1.

4.3. Convergence rate

To establish convergence rates, we add the following assumption.

A5. For $(\alpha, \beta, \theta, \Lambda)$ satisfying A1-A4,

$$\begin{aligned}&\mathrm{E}(l(\alpha, \beta, \theta, \Lambda) - l(\alpha_0, \beta_0, \theta_0, \Lambda_0)) \\ &\qquad \leq -K_1(|\alpha - \alpha_0|^2 + ||\beta - \beta_0||^2 + ||\theta - \theta_0||^2 + m_\Lambda^2(\Lambda, \Lambda_0)),\end{aligned}$$

where K_1 is a fixed positive constant.

Here we assume that the maximizer is "well separated"; this can be verified under the boundedness conditions and differentiability of the log-likelihood function.

Lemma 3. *Under* A1−A5*,*

$$|\hat{\alpha} - \alpha_0|^2 + ||\hat{\beta} - \beta_0||^2 + ||\hat{\theta} - \theta_0||^2 + m_\Lambda^2(\hat{\Lambda}, \Lambda_0) = O_P(n^{-2/3}).$$

We first insert the definition of bracketing number. Let $(\mathbb{F}, ||\cdot||)$ be a subset of a normed space of real functions on some set. Given functions f_1 and f_2, the bracket $[f_1, f_2]$ is the set of all functions f with $f_1 \le f \le f_2$. An ϵ bracket is a bracket $[f_1, f_2]$ with $||f_1 - f_2|| \le \epsilon$. The bracketing number $N_{[]}(\epsilon, \mathbb{F}, ||\cdot||)$ is the minimum number of ϵ brackets needed to cover $\mathbb{F}$. The entropy with bracketing is the logarithm of the bracketing number.

Proof. Lemma 25.84 of van der Vaart (1998) shows that, if A4 is satisfied, there exists a constant K_2 such that for every $\epsilon > 0$, $\log N_{[]}(\epsilon, \{\Lambda\}, L_2) \le K_2(1/\epsilon)$. Since the log-likelihood function is Hellinger differentiable, under A2 and A3 we have $\log N_{[]}(\epsilon, \{l(\alpha, \beta, \theta, \Lambda)\}, L_2) \le K_3(1/\epsilon)$ for a constant K_3.

Apply Theorem 3.2.5 of van der Vaart and Wellner (1996). For $(\alpha, \beta, \theta, \Lambda)$ satisfying $|\alpha - \alpha_0|^2 + ||\beta - \beta_0||^2 + ||\theta - \theta_0||^2 + m_\Lambda^2(\Lambda, \Lambda_0) < \eta$, we have

$$\begin{aligned} &\mathrm{P}^* \sup |\sqrt{n}(\mathrm{E}_n - \mathrm{E})(l(\alpha, \beta, \theta, \Lambda) - l(\alpha_0, \beta_0, \theta_0, \Lambda_0))| \\ &\quad = O_p(1)\eta^{1/2}\left(1 + \frac{\eta^{1/2}}{\eta^2\sqrt{n}}K_4\right) \end{aligned} \tag{4.3}$$

with a constant K_4, where P^* is the outer expectation. According to Theorem 3.2.1 of van der Vaart and Wellner (1996), (4.3) and A5 imply $|\hat{\alpha} - \alpha_0|^2 + ||\hat{\beta} - \beta_0||^2 + ||\hat{\theta} - \theta_0||^2 + m_\Lambda^2(\hat{\Lambda}, \Lambda_0) = O_P(n^{-2/3})$.

We note that the $n^{1/3}$ convergence rate is fundamentally different from the $n^{1/2}$ rate in Kim (2003). Data considered in Kim (2003) have a nonzero proportion of precisely observed event times, i.e, it is a mixture of uncensored data and interval censored data. Thus Kim (2003) is able to achieve the faster convergence rate.

4.4. Information calculation

With semiparametric models, $\sqrt{n}$ consistency and asymptotic normality of estimates of parametric parameters requires non-singularity of the information matrix. We compute the information matrix for (α, β, θ) as follows.

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

$$\dot{l}_\alpha = \left[-\frac{\delta_1 + \delta_2}{1 - p(Z)} + \frac{(1 - \delta_1 - \delta_2)(1 - g_V)}{p(Z) + (1 - p(Z))g_V} \right] \frac{\partial p(Z)}{\partial \alpha},$$
$$\dot{l}_\beta = \left[-\frac{\delta_1 + \delta_2}{1 - p(Z)} + \frac{(1 - \delta_1 - \delta_2)(1 - g_V)}{p(Z) + (1 - p(Z))g_V} \right] \frac{\partial p(Z)}{\partial \beta},$$
$$\dot{l}_\theta = \left\{ \frac{\delta_1 g_U \Lambda(U)}{1 - g_U} - \frac{\delta_2 (g_U \Lambda(U) - g_V \Lambda(V))}{g_U - g_V} - \frac{(1 - \delta_1 - \delta_2) g_V \Lambda(V)}{p(Z) + (1 - p(Z))g_V} \right\} \exp(\theta' X) X.$$

Write $\dot{l}_{\alpha,\beta,\theta} = (\dot{l}_\alpha, \dot{l}'_\beta, \dot{l}'_\theta)'$. Consider a small perturbation of Λ defined by $\Lambda_s = \Lambda + sh$ with $s \sim 0$ and $h \in L_2(P)$, such that Λ_s satisfies A4. We can see that $h = \frac{\partial \Lambda_s}{\partial s}\big|_{s=0}$. Then the score operator for Λ is

$$\begin{aligned} \dot{l}_\Lambda[h] &= \frac{\partial}{\partial s} l(\alpha, \beta, \theta, \Lambda_s)\Big|_{s=0} \\ &= \left(\frac{\delta_1 g_U}{1 - g_U} h(U) - \frac{\delta_2 g_U}{g_U - g_V} h(U) + \frac{\delta_2 g_V}{g_U - g_V} h(V) - \right. \\ &\qquad \left. \frac{(1 - \delta_1 - \delta_2)(1 - p(Z)) g_V}{p(Z) + (1 - p(Z)) G_V} h(V) \right) \exp(\theta' X). \end{aligned}$$

Computing the information matrix requires one to find h^* such that, for any h defined above,

$$\mathrm{E}\{(\dot{l}_{\alpha,\beta,\theta} - \dot{l}_\Lambda[h^*])\dot{l}_\Lambda[h]\} = 0. \tag{4.4}$$

Existence of h^* that satisfies (4.4) can be proved. However, the proof is lengthy and is omitted here. The information matrix is $\mathrm{E}(\dot{l}_{\alpha,\beta,\theta} - \dot{l}_\Lambda[h^*])^{\otimes 2}$, assumed to be positive definite and component-wise bounded.

4.5. Asymptotic normality

We further establish that, despite the slow convergence rate of the cumulative baseline hazard estimate, the estimates of parametric regression coefficients are still $\sqrt{n}$ consistent, asymptotically normally distributed, and efficient (in the sense that any regular estimator would have asymptotic variance equal to or larger than that of the proposed estimate).

Lemma 4. *Under* A1–A5, $\sqrt{n}[(\hat{\alpha}, \hat{\beta}, \hat{\theta}) - (\alpha_0, \beta_0, \theta_0)] \to N(0, E^{-1}(\dot{l}_{\alpha,\beta,\theta} - \dot{l}_\Lambda[h^*])^{\otimes 2})$.

Proof. We list some relevant facts.

1. (Maximization of the objective function) $\mathrm{E}_n \dot{l}_{\alpha,\beta,\theta}(\hat{\alpha},\hat{\beta},\hat{\theta},\hat{\Lambda}) = 0$ component wise; and $\mathrm{E}_n \dot{l}_{\Lambda}[h]|_{\alpha=\hat{\alpha},\beta=\hat{\beta},\theta=\hat{\theta},\Lambda=\hat{\Lambda}} = 0$ for h defined in Section 4.4.

2. (Rate of convergence) $|\hat{\alpha}-\alpha_0|^2 + ||\hat{\beta}-\beta_0||^2 + ||\hat{\theta}-\theta_0||^2 + m_{\Lambda}^2(\hat{\Lambda},\Lambda_0) = O_P(n^{-2/3})$.

3. (Positive Information) The Fisher Information matrix is positive definite and component-wise bounded.

4. (Stochastic equicontinuity) For any $\delta_n \to 0$ and constant $K_5 > 0$, within the neighborhood $\{|\alpha-\alpha_0| < \delta_n, ||\beta-\beta_0|| < \delta_n, ||\theta-\theta_0|| < \delta_n, m(\Lambda,\Lambda_0) < K_5 n^{-1/3}\}$,

$$\sup \sqrt{n}|(\mathrm{E}_n - \mathrm{E})(\dot{l}_{\alpha,\beta,\theta}(\alpha,\beta,\theta,\Lambda) - \dot{l}_{\alpha,\beta,\theta}(\alpha_0,\beta_0,\theta_0,\Lambda_0))| = o_P(1),$$
$$\sup \sqrt{n}|(\mathrm{E}_n - \mathrm{E})(\dot{l}_{\Lambda}[h^*]|_{\alpha,\beta,\theta,\Lambda} - \dot{l}_{\Lambda}[h^*]|_{\alpha_0,\beta_0,\theta_0,\Lambda_0})| = o_P(1).$$

   The above two equations can be established by applying Theorem 3.2.5 of van der Vaart and Wellner (1996) and the entropy result.

5. (Smoothness of the model) Within the neighborhood $\{|\alpha-\alpha_0| < \delta_n, ||\beta-\beta_0|| < \delta_n, ||\theta-\theta_0|| < \delta_n, m(\Lambda,\Lambda_0) < K_5 n^{-1/3}\}$, the expectations of $\dot{l}_{\alpha,\beta,\theta}$ and $\dot{l}_{\Lambda}$ are Hellinger differentiable.

With this in hand, Lemma 4 can be proved using Theorem 3.4 of Huang (1996).

4.6. Inference

In theory, variance of the proposed estimate can be obtained by inverting the information matrix computed in Section 4.4. However we note that the influence function, and hence the information matrix, do not have closed forms, which makes this approach very difficult.

The key to establishing the validity of the weighted bootstrap is that the functional set $\{w \times l(\alpha,\beta,\theta,\Lambda)\}$ has the same entropy and similar asymptotic behaviors as the set $\{l(\alpha,\beta,\theta,\Lambda)\}$. The unconditional properties of the weighted MLE defined in Section 3.4 can be established following the same arguments as for the ordinary MLE, which implies that, conditional on the observed data, $(\hat{\alpha}^*,\hat{\beta}^*,\hat{\theta}^*) - (\hat{\alpha},\hat{\beta},\hat{\theta})$ has the same asymptotic variance as $(\hat{\alpha},\hat{\beta},\hat{\theta}) - (\alpha_0,\beta_0,\theta_0)$. We refer to Ma and Kosorok (2005b) for more details.

5. Simulation Study

We conducted simulations to investigate finite sample performance of the proposed estimator. Here, we considered two covariates $Z = (Z_1, Z_2)$ and set

$X = Z$. The cure indicator and event time (if not cured) were generated from models (2.1) and (2.2), with $(\alpha_0, \beta_0, \theta_0) = (-2, 2, -2, 2, -1)$. We generated the censoring times $\tilde{U}_j = \sum_{i=1}^{j} \xi_i$, where $\xi_i \sim Unif[0.1, 0.25]$. We kept generating ξ_i until $\tilde{U}_j > min(T, 2.5)$. We considered the following three simulation scenarios.

1. With probabilities 0.5, $Z_1 = 0.5$ or 1.5, while Z_2 had the same distribution as Z_1. On average, about 41% observations were left censored, 31% were interval censored, 9% were right censored and not cured, and the rest were cured.

2. With probabilities 0.5, $Z_1 = 0.5$ or 1.5, and $Z_2 \sim Unif[0, 2]$; censoring rates were similar to those under Scenario 1.

3. $Z_1, Z_2 \sim Unif[0, 2]$; censoring rates were similar to those under Scenario 1.

Under Scenarios 1–3, we considered covariates with discrete, mixed, and continuous distributions, respectively. We considered sample sizes 200, 400, and 800. Summary statistics based on 500 replicates are shown in Table 1. Our simulation study suggests that (1) the proposed estimates have very small biases, even for sample size as small as 200; (2) standard deviations of the estimates shrink at approximately the $\sqrt{n}$ rate, which partly supports Lemma 4; (3) the weighted bootstrap estimated standard deviations are very close to those of the original estimates, which supports validity of the weighted bootstrap. More simulations under different settings showed similar, satisfactory results.

6. Analysis of the HDSD

When analyzing the HDSD, one record with missing measurements is removed and 238 records are available for downstream analysis. The response of interest is the time to onset of grade IV VGE, which is interval censored. Covariates of interest include age, sex, TR360, and NOADYN. Age has mean 31.882, standard deviation 7.126, and range (20, 54); to make covariates more comparable, we divided age by 10. Of 238 subjects, 177 were male; in the model, the female group is used as reference. TR360 is a measure of decompression stress, the ratio of the partial pressure of nitrogen to ambient pressure at the final altitude, and has mean 1.637, standard deviation 0.227, and range (1.040, 1.890). NOADYN is an experimentally manipulated variable and an indicator for whether the test subject was ambulatory (NOADYN=1) or lower body adynamic (NOADYN=0) during the test session; 195 records have NOADYN=1. We refer to Thompson and Chhikara (2003) for more information on the experiments and covariates.

We analyze the HDSD with the proposed two-part model. Considering that "only individual characteristics can influence an individual's susceptibility to

Table 1. Simulation study: means computed based on 500 replicates. *est*: estimate; *sd*: standard deviation of $(\hat{\alpha}, \hat{\beta}, \hat{\theta})$; $\hat{sd}$: standard deviation of $(\hat{\alpha}^*, \hat{\beta}^*, \hat{\theta}^*) - (\hat{\alpha}, \hat{\beta}, \hat{\theta})$.

			α	β_1	β_2	θ_1	θ_2
Scenario 1	$n = 200$	*est*	-2.001	1.904	-1.993	1.954	-0.988
		sd	0.984	0.952	0.749	0.323	0.264
		$\hat{sd}$	0.935	0.922	0.763	0.374	0.271
	$n = 400$	*est*	-2.019	2.102	-2.028	1.967	-0.976
		sd	0.678	0.645	0.506	0.240	0.196
		$\hat{sd}$	0.661	0.712	0.500	0.248	0.202
	$n = 800$	*est*	-2.068	2.066	-2.034	1.995	-0.984
		sd	0.480	0.452	0.372	0.173	0.148
		$\hat{sd}$	0.486	0.439	0.347	0.175	0.147
Scenario 2	$n = 200$	*est*	-2.032	1.963	-2.006	1.926	-1.010
		sd	1.195	0.943	0.596	0.294	0.253
		$\hat{sd}$	1.229	0.901	0.600	0.286	0.246
	$n = 400$	*est*	-2.062	1.984	-2.025	1.986	-0.945
		sd	0.859	0.712	0.414	0.197	0.170
		$\hat{sd}$	0.866	0.700	0.408	0.204	0.164
	$n = 800$	*est*	-1.949	2.026	-1.940	2.004	-0.947
		sd	0.597	0.509	0.307	0.136	0.123
		$\hat{sd}$	0.606	0.485	0.306	0.136	0.116
Scenario 3	$n = 200$	*est*	-2.121	2.002	-1.989	1.992	-0.980
		sd	1.134	0.884	0.824	0.276	0.286
		$\hat{sd}$	1.165	0.865	0.867	0.279	0.239
	$n = 400$	*est*	-2.095	1.937	-1.921	1.905	-0.950
		sd	0.874	0.546	0.547	0.173	0.201
		$\hat{sd}$	0.802	0.567	0.615	0.186	0.198
	$n = 800$	*est*	-2.008	2.026	-2.079	1.932	-0.938
		sd	0.522	0.415	0.402	0.114	0.136
		$\hat{sd}$	0.484	0.428	0.468	0.104	0.137

grade IV VGE" (Thompson and Chhikara (2003)), we set $Z = (age,\ sex)$. All four covariates are included in the Cox model.

We compute the MLE using the iterative algorithm described in Section 3.3. Satisfactory convergence is achieved. The variance estimation is obtained using the weighted bootstrap and exp(1) weights. The MLEs are

$$\hat{\alpha} = 7.393(2.017),$$
$$\hat{\beta}_1 = -1.741(0.593), \hat{\beta}_2 = -2.522(0.927),$$
$$\hat{\theta}_1 = -0.602(0.188), \hat{\theta}_2 = -0.702(0.710), \hat{\theta}_3 = 0.561(0.413), \hat{\theta}_4 = 1.021(0.458),$$

where values in the "()" are the estimated standard errors.

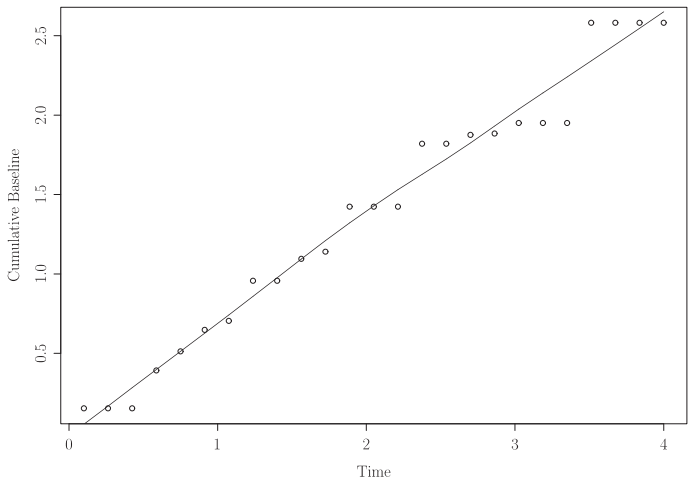

Figure 7.1. Analysis of HDSD: estimated cumulative baseline hazard function.

From the above estimation results, we conclude that both age and sex have significant effects on the risk of getting grade IV VGE; specifically, older and/or male are more susceptible, a finding consistent with Thompson and Chhikara (2003). We find for susceptible subjects that the effects of age and NOADYN are significant; specially, younger subjects and/or subjects being ambulatory (NOADYN=1) experience grade IV VGE faster, a contrast with Thompson and Chhikara (2003) who do not conclude significance of the age effect. We find the effects of sex and TR360 on survival are not significant, consistent with Thompson and Chhikara (2003).

In Figure 7.1, we show the estimated cumulative baseline hazard. We also provide the lowess smoother, which suggests that the cumulative baseline is close to a linear function. Thus, it may be possible to assume constant hazard and simplify the proposed model. Since significantly different techniques are involved, we do not pursue this.

7. Discussion

Although the data structure investigated in this article is specific (i.e., there exists a cured subgroup), we expect that the proposed methodology can be extended to other heterogeneous data with minor modifications. We have assumed the parametric generalized linear model and Cox model. We expect that the estimation and inference approaches, and their asymptotic properties, can be extended to other models.

The HDSD dataset analyzed in Section 6 is in fact a subset of the data in Thompson and Chhikara (2003). In the entire HDSD study, there were multiple

experiments, whereas we focused on only one. Experiments following the first were conducted on a subset of subjects, where subjects *volunteered* (instead of being randomly selected). This raises serious concerns on the possibility of biased sampling. For example, even with the same experimental scheme, the first experiment had a censoring rate of 71%, whereas the experiments that followed had a censoring rate of 83%. Significant differences in covariates also exist. Investigating the possibility of biased sampling is interesting, but beyond the scope of this article.

Acknowledgement

The author would like to thank the Editor and three reviewers for their careful review and insightful comments that have led to significant improvement of this article. This study has been partly supported by DMS 0805984 from NSF.

References

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

Conkin, J., Bedahl, S. and Van Liew, H. (1992). A computerized databank of decompression sickness incidence in altitude chambers. *Aviation, Space and Enviromental Medicine* **63**, 819-824.

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

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

Jin, J., Ying, Z. and Wei, L. J. (2001). A simple resampling method by perturbing the minimand. *Biometrika* **88**, 381-390.

Kim, J. S. (2003). Maximum likelihood estimation for the proportional hazards model with partly interval-censored data. *J. Roy. Statist. Soc. Ser. B* **65**, 489-502.

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. (2009). Cure model with current status data. *Statist. Sinica* **19**, 233-249.

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.

Robertson, T., Wright, F. T. and Dykstra, R. L. (1988). *Order Restricted Statistical Inference.* Wiley.

Schick, A. and Yu, Q. (2000). Consistency of the GMLE with mixed case interval-censored data. *Scandinavian Journal of Statistics* **27**, 45-55.

Sen, B. and Banerjee, M. (2007). A pseudolikelihood method for analyzing interval censored data. *Biometrika* **94**, 71-86.

Song, S. (2004). Estimation with univariate "mixed case" interval censored data. *Statist. Sinica* **14**, 269-282.

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 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.

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.

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

E-mail: shuangge.ma@yale.edu

(Received July 2008; accepted February 2009)