

Additive hazards models with latent treatment effectiveness lag time

BY YING QING CHEN*

*Division of Biostatistics, School of Public Health,
University of California, Berkeley, CA 94720, U.S.A.*

CHARLES A. ROHDE, MEI-CHENG WANG

*Department of Biostatistics, School of Hygiene and Public Health
Johns Hopkins University, Baltimore, MD 21205, U.S.A.*

* Send correspondence to: Ying Qing Chen, Division of Biostatistics, School of Public Health, 140 Warren Hall #7360, University of California, Berkeley, CA 94720. E-mail: yqchen@stat.berkeley.edu. Tel: (510) 642-7900. Fax: (510) 643-5163.

Additive hazards models with latent treatment effectiveness lag time

BY YING QING CHEN

Division of Biostatistics, School of Public Health

University of California, Berkeley, CA 94720, U.S.A.

CHARLES A. ROHDE AND MEI-CHENG WANG

Department of Biostatistics, School of Hygiene and Public Health

Johns Hopkins University, Baltimore, MD 21205, U.S.A.

SUMMARY

In many clinical trials to evaluate treatment efficacy, it is believed that there may exist latent treatment effectiveness lag times after which medical treatment procedure or chemical compound would be in full effect. In this article, semiparametric regression models are proposed and studied to estimate the treatment effect accounting for such latent lag times. The new models take advantage of the invariance property of the additive hazards model in marginalizing over random effects, so parameters in the models are easy to be estimated and interpreted, while the flexibility without specifying baseline hazard function is kept. Monte Carlo simulation studies demonstrate the appropriateness of the proposed semiparametric estimation procedure. Data collected in the actual randomized clinical trial, which evaluates the effectiveness of biodegradable carmustine polymers for treatment of recurrent brain tumors, are analyzed.

Some key words: Change Point; Clinical trials; Cure models; Mixture models; Random effects; Semiparametric model; Survival data.

1 Introduction

In comparative randomised clinical trials, efficacy of a new treatment, e.g., a new drug or medical procedure, is often assessed by comparing collected survival data. As expected, not all proposed treatments are always taking effect as soon as their initiation. In fact, many treatments are observed to have slow onset of action after their initiation, such as in Pérez et al. (1997) to assess the efficacy of an antidepressant treatment. For such phenomenon of slow onset of action, researchers often believe that there may exist a so-called treatment effectiveness lag time before the treatment becomes fully effective (Wu, Fisher & DeMets, 1980; Gail, 1985; Lakatos, 1986; Zucker & Lakatos, 1990). A treatment effectiveness lag time is the time for a biological subject to fully respond to medical procedures or compounds. It is usually not observable, although certain biomarkers can be used to artificially define the termination of treatment effectiveness lag time. If the treatment effectiveness lag times are ignored, the assumptions of the widely-used proportional hazards model (Cox, 1972) with constant proportionality are often inappropriate and hence not able to correctly accommodate the observation of slow onset of action, unless some ad hoc time-dependent structure is included.

Although the existence of treatment effectiveness lag time was recognized, in most of previous research in developing appropriate statistical methodologies, researchers used the notion of a common fixed treatment effectiveness lag time for every individual or tried to find ad hoc time-dependent lag functions for the proportional hazards model, for instance, in Self et al. (1988) and Zucker & Lakatos (1990). But due to the heterogeneity among the biological subjects, such as unobservable different genotypes, the treatment effectiveness lag times could apparently vary individual-by-individual. In addition, since prior knowledge about the lag is often rarely available, an accurate lag function is usually unknown and

difficult to be determined.

To account for the latent treatment effectiveness lag time and its heterogeneity among individuals, an unobservable random variable, U , say, is introduced in this article to represent such lag time, which is treated as a random effect. In addition, since some of the treatment effectiveness lag times are too long to allow the full onset of action, a mixture cure model (Farewell, 1982; Gray & Tsiatis, 1989; Laska & Meisner, 1992) will be adapted for U . Furthermore, to identify the subject-dependent proportion of long-term treatment effectiveness lag times, appropriate regression models will be incorporated into the mixture cure model.

One straightforward approach to estimate the treatment effect accounting for the latent U is through the proportional hazards model. For a specific example, given a treatment effectiveness lag time $U = u > 0$, the relative hazards ratio can be assumed as 1 before u and β after u . This is in fact the simplest version of the proportional hazards model with change point as random effect (Nguyen, Rogers & Walker, 1984; Basu, Gosh & Joshi, 1988). Although it carries simple form and straightforward interpretation conditioning on the random effect, its marginalized version over the random effect does not own clean multiplicative form any longer. This leads to some serious consequences, such as “numerical and theoretical difficulties” in inference procedures and “awkward interpretation” in parameters, as pointed out in Lin & Ying (1997).

Instead of the multiplicative proportional hazards model, we will propose and study the change point hazards models with additive random effects to determine the covariate effect. The remainder of this article is organized as follows. In §2, we will present the mixture model. The semiparametric inference procedures and its asymptotic properties are studied in §3. Numerical studies are demonstrated in §4. Some concluding remarks and discussion

are in §5. Mathematical proofs are collected in Appendices.

2 The mixture model

Suppose that there are n independent participants in the study. For $i = 1, 2, \dots, n$, the failure time and censoring time for individual i are T_i and C_i , respectively; and U_i is the latent treatment effectiveness lag time, i.e., after which the treatment is fully effective. The actual observed data consist of the triplets of (X_i, Δ_i, Z_i) . Here $X_i = \min(T_i, C_i)$ is the survival time, and

$$\Delta_i = I(T_i \leq C_i) = \begin{cases} 1 & \text{if } T_i \leq C_i \\ 0 & \text{otherwise} \end{cases}$$

is the censoring indicator, where $I(\cdot)$ is the indicator function taking the value of 1 if the condition is satisfied and 0 otherwise. Let superscript T denote the transpose of vector or matrix and $Z_i(t) = (W_i^T(t), R_i^T(t))^T$ be the p -vector covariate. In particular, to estimate a treatment effect such as in a two-arm randomised clinical trial, W_i can be the treatment indicator being 1 if the participant is in the treatment group and 0 otherwise; and $R_i(t)$ is the concomitant risk factors or confounding variables, for which the treatment effect needs to be adjusted, such as demographic variables or social-economic status. In addition, we assume that (T_i, C_i) are independent conditional on Z_i .

2.1 Distribution of treatment effectiveness lag times

In practice, it is noticeable that there exists possibility with which a portion of population may never respond to the treatment. For example, when the treatment dosage does not meet the participant's minimal threshold for response, the treatment may never be able to take full effect. In this case, the treatment effectiveness lag is considered as "long-term." Or, when the treatment effectiveness lag time is relatively long enough to exceed certain

pre-determined time point u_0 , e.g., 6 weeks in antidepressant therapy trials (Pérez et al., 1997), the treatment effectiveness lag time is also considered as “long-term”. Otherwise, the treatment effectiveness lag time subject to early full treatment response is called “short-term.”

Denote Y_i the indicator of short-term treatment effectiveness lag time for the i th participant, i.e.

$$Y_i = \begin{cases} 1 & \text{if } i\text{th treatment effectiveness lag time is short-term;} \\ 0 & \text{if } i\text{th treatment effectiveness lag time is long-term.} \end{cases}$$

Furthermore, let $\bar{F}_0(t; \tau) = 1 - F_0(t; \tau)$ be the conditional survival function for $Y_i = 1$, $i = 1, 2, \dots, n$. Then the treatment effectiveness lag time U_i 's survival function $\bar{G}_i(t) = 1 - G_i(t)$, $t \in [0, \infty)$ is assumed of the cure mixture model (Farewell, 1992):

$$\begin{aligned} \bar{G}_i(t) &= \Pr\{Y_i = 1\} \times \Pr\{U_i \geq t | Y_i = 1\} + \Pr\{Y_i = 0\} \\ &= p_i \bar{F}_0(t; \tau_0) + (1 - p_i) \end{aligned} \tag{1}$$

for $i = 1, 2, \dots, n$, where $p_i \in [0, 1]$. There are varieties of choices for $F(t; \tau)$, for example, distributions of Exponential, Weibull and Gamma.

From model (1), it seems in form that $G_i(t)$ is not a rigorously defined distribution function in probability theory, whenever $p_i < 1$. However, it implicitly carries the message that a treatment effectiveness lag time can be long-term, or even infinite, which exactly describes the possible scenarios discussed above. If necessary, to make G_i more statistically concrete, for example, an artificial truncation time, u_0 , say, can be chosen. Then the form of $\bar{G}_i(t)$ does not change when $0 \leq t \leq u_0$ but is 0 when $t > u_0$ (Laska & Meisner, 1992; Tamura, Faries & Feng, 2000). Nevertheless, whether or not choosing a truncation time should not undermine the development of our proposed method in this article, as seen in the later development.

In model (1), p_i is the probability of the i th participant having short-term treatment effectiveness lag time. The larger the magnitude of p_i , the easier the treatment to be fully effective within a reasonable time range. It can also be linked to the corresponding covariate Z_i through appropriate regression models. For example, the logistic regression models (Farewell, 1992) can be used:

$$\log \frac{p_i(\alpha_0)}{1 - p_i(\alpha_0)} = \alpha_0^T Z_i(0). \quad (2)$$

Other choices include probit, log-log and complementary log-log regression models (McCullagh & Nelder, 1989, p. 108).

2.2 Additive hazards models with latent lag time

Denote $\lambda(\cdot)$ as the hazard function for the failure time, T . We first use the following model to determine the covariate effect with the treatment effectiveness lag time:

$$\lambda\{t|Z_i(t), U_i; \theta_0\} = \lambda_0(t) + \gamma_0^T R_i(t) + I(U_i \leq t)\beta_0^T W_i(t), \quad (3)$$

where $\theta_0 = (\beta_0^T, \gamma_0^T)^T$ is p -vector parameter and $\lambda_0(t)$ an unknown baseline hazard function. In model (3), conditional on the treatment effectiveness lag time U_i , the hazard function of $Z_i(t)$ is $\lambda_0(t) + \gamma_0^T R_i(t)$ before U_i and $\lambda_0(t) + \gamma_0^T R_i(t) + \beta_0^T W_i(t)$ after. Therefore, the parameter β_0 characterizes the full effect of $W_i(t)$ after the treatment effectiveness lag time, which is often of the most interest, e.g., when $W_i(t)$ is the treatment indicator.

Model (3) is a change point model which generalizes the notion of fixed treatment effectiveness lag time in Zucker & Lakatos (1990) by introducing heterogeneous treatment effectiveness lag time U_i 's. This reflects the truth that different individuals may have different paces and hence different treatment effectiveness lag times. Jointly, model (1) and (3) determine both the probability with which the treatment effect is fully effective within a reasonable time range and the magnitude of full treatment effect.

Model (3) is also an additive hazards model with random effects. As argued in Breslow & Day (1980, p. 53-59; 1987, p. 122-131) and Lin & Ying (1997, p. 188-189), an additive hazards model is able to provide sound interpretation in clinical studies. And, more importantly, it yields a much simpler marginal model after the random effects are integrated. As shown in Appendix 1, the marginalized model (3) is:

$$\lambda\{t|Z_i(t), \theta_0\} = \lambda_0(t) + \gamma_0^T R_i(t) + \beta_0^T W_i(t) H_i(t; \beta_0, \phi_0), \quad (4)$$

where

$$H_i(t; \beta_0, \phi_0) = \frac{\int_0^t e^{-\beta_0^T \int_u^t W_i(s) ds} dG(u; \phi_0)}{\int_0^T e^{-\beta_0^T \int_u^t W_i(s) ds} dG(u; \phi_0) + \bar{G}(t; \phi_0)}, \quad (5)$$

and $\phi_0 = (\tau_0, \alpha_0)$.

Assume that $R_i(t)$ and $W_i(t)$ are bounded. Furthermore, if there exists an $i_0 \in \{1, 2, \dots, n\}$ such that $\Pr\{U_{i_0} \leq T_{i_0}\} > 0$, then it is easily seen that when $p = 1$,

1. $0 \leq H(t; \beta_0, \phi_0) \leq 1$;
2. $\lim_{t \rightarrow 0} H(t; \beta_0, \phi_0) = 0$, $\lim_{t \rightarrow \infty} H(t; \beta_0, \phi_0) = 1$;
3. $H(t; \beta_0, \phi_0)$ is non-decreasing.

Here, the additional assumption of $\Pr\{U_{i_0} \leq T_{i_0}\} > 0$ is an identifiability condition to secure the estimability of parameter β_0 and the above properties. Otherwise, $E[I(U_i \leq T_i) \beta_0^T W_i(t)] \equiv 0$ for any i and hence β_0 is not estimable. This essentially requires that the treatment effectiveness lag time is not always longer than the failure time for every individual in study.

In fact, $H(t; \beta_0, \phi_0)$ corresponds to the lag function the researchers have been looking for. It is also of interest that $H(t; \beta_0, \phi_0)$ owns similar properties of cumulative distribution function (CDF). When $W(t) \equiv 0$, $H(t)$ is exactly the distribution function of G , although

its effect on the hazard function is nullified by zero $W(t)$. In §4, we will demonstrate what $H(t; \beta_0, \phi_0)$ may appear to be by studying some special $H(t; \beta_0, \phi_0)$'s.

In general, the identifiability can be critical for models with arbitrary random effects. This usually does not impose serious challenges upon model (3), as seen in the following theorem:

THEOREM 1. *If $W_i(t)$ and the parameter space \mathcal{B} are bounded, then model (3) is identifiable if and only if ϕ in G is identifiable.*

To prove the above theorem, it is sufficient to show that for two treatment effectiveness lag time distribution functions of G_1 and G_2 , they are equal almost everywhere if

$$\int_0^\infty e^{-\beta_0^\top I(u \leq t) \int_u^t W_i(s) ds} dG_1(u; \phi_0) = \int_0^\infty e^{-\beta_0^\top I(u \leq t) \int_u^t W_i(s) ds} dG_2(u; \phi_0),$$

which is straightforward to be established under the assumed conditions.

3 Inference Procedures and Asymptotic Properties

In this section, we present a semiparametric estimation procedure by fully utilizing the unaltered additive structure in model (3).

Let $N_i(t) = I(X_i \leq t, \Delta_i = 1)$ and $Y_i(t) = I(X_i \geq t)$, $i = 1, 2, \dots, n$. Consider the filtration \mathcal{F}_t defined by

$$\mathcal{F}_t = \sigma\{N_i(t), Y_i(t), Z_i(t); i = 1, 2, \dots, n\}.$$

Define

$$M_i(t; \theta, \phi, \Lambda_0) = P_i(t; \theta, \phi) - \int_0^t Y_i(s) d\Lambda_0(s),$$

where $P_i(t; \theta, \phi) = N_i(t) - \int_0^t Y_i(s) \{\gamma^\top R_i(s) + \beta^\top W_i(s) H_i(s; \theta, \phi)\} ds$. It is true that $M_i(\cdot; \theta_0, \phi_0)$ are local square integrable martingales of \mathcal{F}_t . Therefore similar to the partial score equations

for the proportional hazards model (Fleming & Harrington, 1991), the following estimating equations can be used to estimate (θ_0, ϕ_0) :

$$\sum_{i=1}^n \int_0^{\infty} Q(t; \theta, \phi) J_i(t; \theta, \phi) dM_i(t; \theta, \phi, \Lambda_0) = 0, \quad (6)$$

where $Q(t; \theta_0, \phi_0)$ is a measurable weight function with respect to \mathcal{F}_t , which converges uniformly to a deterministic function of $q(t; \theta_0, \phi_0)$, and $J_i(t; \theta, \phi)$ are smooth functions of same dimension of (θ, ϕ) , which are also predictable processes of $t, i = 1, 2, \dots, n$.

Although the baseline hazard function is unknown in (6), a reasonable estimator of $\Lambda_0(t)$ of Breslow-type, nevertheless, is

$$\hat{\Lambda}_0(t; \theta, \phi) = \int_0^t \left\{ \sum_{i=1}^n dP_i(s; \theta, \phi) \right\} \left\{ \sum_{i=1}^n Y_i(s) \right\}^{-1} \quad (7)$$

as in Lin and Ying (1994). Thus we can use the following equations to estimate the parameters of interest:

$$\sum_{i=1}^n \int_0^{\infty} Q(t; \theta, \phi) J_i(t; \theta, \phi) d\hat{M}_i(t; \theta, \phi, \hat{\Lambda}_0) = 0. \quad (8)$$

where $\hat{M}_i(t; \theta, \phi, \hat{\Lambda}_0) = P_i(t; \theta, \phi) - \int_0^t Y_i(s) d\hat{\Lambda}_0(s; \theta, \phi)$.

Denote the left-hand side of equation (6) as $\Gamma(\theta, \phi)$. Some algebraic manipulation shows that $\Gamma(\theta, \phi)$ is equal to

$$\Gamma(\theta, \phi) = \sum_{i=1}^n \int_0^{\infty} Q(t; \theta, \phi) \{J_i(t; \theta, \phi) - \bar{J}(t; \theta, \phi)\} dP_i(t; \theta, \phi) \quad (9)$$

where $\bar{J}(t; \theta, \phi) = \{\sum_{i=1}^n Y_i(t) J_i(t; \theta, \phi)\} \{\sum_{i=1}^n Y_i(t)\}^{-1}$.

To study the asymptotic properties of solutions by solving $\Gamma(\theta, \phi) = 0$, we first assume the following regularity conditions:

1. There exists a time $t_0 > 0$ such that $\lim_{n \rightarrow \infty} \sum_{i=1}^n Y_i(t_0) > 0$;

2. There exists an integrable function $v(t)$ such that, for any $t \in [0, t_0]$,

$$n^{-1} \sum_{i=1}^n Y_i(t) \{J_i(t; \theta_0, \phi_0) - \bar{J}(t; \theta_0, \phi_0)\}^{\otimes 2} \lambda_i(t|Z_i) - v(t) \xrightarrow{P} 0.$$

where $a^{\otimes 0} = 1$, $a^{\otimes 1} = a$ and $a^{\otimes 2} = aa^T$;

3. For any $\epsilon > 0$ such that

$$\begin{aligned} & n^{-1} \sum_{i=1}^n \int_0^{t_0} Y_i(s) \lambda_i(t|Z_i) \\ & \times \|J_i(t; \theta_0, \phi_0) - \bar{J}(t; \theta_0, \phi_0)\|^2 I\{n^{-1} \|J_i(t; \theta_0, \phi_0) - \bar{J}(t; \theta_0, \phi_0)\|^2 > \epsilon\} ds \xrightarrow{P} 0, \end{aligned}$$

where $\|\cdot\|$ defines the Euclidean norm.

Consider the process $\Gamma(t; \theta, \phi) = \sum_{i=1}^n \int_0^t Q(s; \theta, \phi) \{J_i(s; \theta, \phi) - \bar{J}(s; \theta, \phi)\} dP_i(s; \theta, \phi)$. It is also true that $\Gamma(t; \theta_0, \phi_0)$ is an \mathcal{F}_t -martingale.

LEMMA 2. *With regularity conditions 1-3, $n^{-1/2}\Gamma(t; \theta_0, \phi_0)$ converges weakly in $\mathfrak{D}[0, t_0]$ to a zero-mean Gaussian process with independent increments and the variance function $V(t; \theta_0, \phi_0) = \int_0^t q(s)v(s)ds$, which is*

$$\lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n \int_0^t Q(s; \theta_0, \phi_0) Y_i(s) \{J_i(s; \theta_0, \phi_0) - \bar{J}(s; \theta_0, \phi_0)\}^{\otimes 2} \lambda_i(s|Z_i) ds.$$

PROOF. *See Appendix 2.1.*

Define the solution of $\Gamma(\theta, \phi) = 0$ as $(\hat{\theta}, \hat{\phi})$. The following theorem establishes the asymptotic properties for $(\hat{\theta}, \hat{\phi})$.

THEOREM 3. *Suppose that there exists non-singular $D(\theta_0, \phi_0)$ such that*

$$D(\theta_0, \phi_0) = \lim_{n \rightarrow \infty} -n^{-1} \begin{pmatrix} \partial \Gamma(\theta_0, \phi_0) / \partial \theta^T \\ \partial \Gamma(\theta_0, \phi_0) / \partial \phi^T \end{pmatrix}.$$

With the regularity conditions 1-3, if all the partial derivatives are bounded and continuous in a neighbourhood of (θ_0, ϕ_0) , then $(\hat{\theta}, \hat{\phi})$ is uniquely defined and

$$n^{1/2} \begin{pmatrix} \hat{\theta} - \theta_0 \\ \hat{\phi} - \phi_0 \end{pmatrix} \rightarrow N\{0, D^{-1}(\theta_0, \phi_0) V(\theta_0, \phi_0) D^{-1}(\theta_0, \phi_0)^T\}, \quad (10)$$

where $V(\theta_0, \phi_0) = V(t_0; \theta_0, \phi_0)$.

PROOF. See Appendix 2.2.

In practice, to make inference about the estimates of parameters, it is natural to use the empirical estimates of its asymptotic variance-covariance matrix by its consistent estimator $\hat{D}^{-1}(\hat{\theta}, \hat{\phi})\hat{V}(\hat{\theta}, \hat{\phi})\hat{D}^{-1}(\hat{\theta}, \hat{\phi})^T$, where

$$\begin{aligned}\hat{D}(\theta, \phi) &= n^{-1} \sum_{i=1}^n \int_0^{t_0} Q(t, \theta, \phi) Y_i(t) \{J_i(t; \theta, \phi) - \bar{J}(t; \theta, \phi)\} \{\gamma^T R_i(t) + \beta^T W_i(t) H_i(t; \theta, \phi)\}' dt \\ \hat{V}(\hat{\theta}, \hat{\phi}) &= n^{-1} \sum_{i=1}^n \int_0^{t_0} Q(t, \theta, \phi) \{J_i(t; \theta, \phi) - \bar{J}(t; \theta, \phi)\}^{\otimes 2} dN_i(t).\end{aligned}$$

Furthermore, replacing the parameters of (θ, ϕ) with $(\hat{\theta}, \hat{\phi})$ in (7) leads to a natural estimator of the cumulative baseline hazard function. As shown in Appendix 2.3, $n^{1/2}\{\hat{\Lambda}_0(t; \hat{\theta}, \hat{\phi}) - \Lambda_0(t)\}$ converges weakly to a zero-mean Gaussian process with covariance function of $\Sigma(t_1, t_2)$, which is the limit of

$$\begin{aligned}& \int_0^{\min(t_1, t_2)} \frac{n}{\sum_{i=1}^n Y_i(t)} \frac{\sum_{i=1}^n dN_i(t)}{\sum_{i=1}^n Y_i(t)} \\ & + K_1^T(t_2) D^{-1} V (D^{-1})^T K_1(t_1) - K_1^T(t_2) D^{-1} K_1(t_1) - K_1^T(t_1) D^{-1} K_2(t_2).\end{aligned}$$

where

$$\begin{aligned}K_1(t; \theta_0, \phi_0) &= \int_0^t \left[\sum_{i=1}^n Y_i(s) \{\gamma_0^T R_i(s) + \beta_0^T W_i(s) H_i(s; \theta_0, \phi_0)\}' \right] \left\{ \sum_{i=1}^n Y_i(s) \right\}^{-1} ds, \\ K_2(t; \theta_0, \phi_0) &= \int_0^t \left(\left[\sum_{i=1}^n Y_i(s) \{\gamma_0^T R_i(s) + \beta_0^T W_i(s) H_i(s; \theta_0, \phi_0)\} J_i(s; \theta_0, \phi_0) \right] \left\{ \sum_{i=1}^n Y_i(s) \right\} \right. \\ & \quad \left. - \left[\sum_{i=1}^n Y_i(s) \{\gamma_0^T R_i(s) + \beta_0^T W_i(s) H_i(s; \theta_0, \phi_0)\} \right] \left\{ \sum_{i=1}^n Y_i(s) J_i(s; \theta_0, \phi_0) \right\} \right) \\ & \quad \times \left\{ \sum_{i=1}^n Y_i(s) \right\}^{-2} ds.\end{aligned}$$

Although the proposed estimating equations can be viewed as parallel to the partial score equations for the proportional hazards model, they are still ad hoc. However, by the

techniques in Lai & Ying (1992) and Lin & Ying (1994), we can compute the semiparametric efficiency bound for the family of parametric submodels as

$$\lambda\{t|Z(t)\} = \lambda_0(t) + \gamma^T R(t) + \beta^T W(t)H(t; \beta, \phi) + \psi\eta(t),$$

where γ , β , ϕ and ψ are parameters and $\eta(\cdot)$ is a fixed function. As a result, the optimal estimating function for θ_0 is computed as:

$$\Gamma_{\text{opt}}(\theta, \phi) = \sum_{i=1}^n \int_0^{\infty} \{\lambda_0(t) + \gamma^T R_i(t) + \beta^T W_i(t)H_i(t; \beta, \phi)\}^{-1} \{J_i^*(t; \theta, \phi) - \bar{J}^*(t; \theta, \phi, \lambda_0)\} dP_i(t; \theta, \phi),$$

where

$$J_i^*(t; \theta, \phi) = \begin{pmatrix} R_i(t) \\ W_i(t)H_i(t) + \beta^T W_i(t)H'_\beta(t; \beta, \phi) \\ \beta^T W_i(t)H'_\phi(t; \beta, \phi) \end{pmatrix},$$

and

$$\bar{J}^*(t; \theta, \phi, \lambda) = \frac{\sum_{i=1}^n Y_i(t) \{\lambda_0(t) + \gamma^T R_i(t) + \beta^T W_i(t)H_i(t; \beta, \phi)\}^{-1} J_i^*(t; \theta, \phi)}{\sum_{i=1}^n Y_i(t) \{\lambda_0(t) + \gamma^T R_i(t) + \beta^T W_i(t)H_i(t; \beta, \phi)\}^{-1}}.$$

However it is difficult to use Γ_{opt} in practice, because the estimating functions themselves involve the baseline hazard function. Although adaptive procedures using special techniques such as sample-splitting in Lin & Ying (1994) are available, the estimation of λ_0 always imposes an imminent challenge, especially when sample size is small.

To practically implement the optimal estimating functions, similar versions can be used instead for convenience. For example, one choice suggested in Lin & Ying (1994, 1995) is to use Γ_{opt} without including $\{\lambda_0(t) + \gamma^T R_i(t) + \beta^T W_i(t)H_i(t; \beta, \phi)\}^{-1}$, i.e.,

$$\Gamma^*(\theta, \phi) = \sum_{i=1}^n \int_0^{\infty} \{J_i^*(t; \theta, \phi) - \tilde{J}^*(t; \theta, \phi, \lambda_0)\} dP_i(t; \theta, \phi), \quad (11)$$

where

$$\tilde{J}^*(t; \theta, \phi, \lambda_0) = \frac{\sum_{i=1}^n Y_i(t) J_i^*(t; \theta, \phi)}{\sum_{i=1}^n Y_i(t)}.$$

Apparently, when the ignored term does not vary much from a constant, then Γ^* should not lose much efficiency.

4 Numerical Studies

4.1 Examples of Lag Function $H(t; \beta_0, \phi_0)$

As proposed in section 2.1, the distribution of G is a mixture cure distribution. In this section, we select some special G 's to demonstrate their corresponding lag functions.

As seen in the definition of G , $\bar{F}_0(t)$ is in fact the conditional probability that the treatment effectiveness lag time occurs after time t for short-term treatment effectiveness lag times. One simple example is that let $F_0(t)$ be Exponential with survival function of:

$$\bar{F}_0(t; \tau_0) = e^{-\tau_0 t} I(t \geq 0).$$

Furthermore, let $W(t)$ be constant W_0 . Then

$$\begin{aligned} H(t; \beta_0, \phi_0) &= \frac{\tau_0 p(\alpha_0)}{\beta_0 W_0 - \tau_0} (e^{-\tau_0 t} - e^{-\beta_0 W_0 t}) \\ &\times \left[\frac{\tau_0 p(\alpha_0)}{\beta_0 W_0 - \tau_0} (e^{-\tau_0 t} - e^{-\beta_0 W_0 t}) + p(\alpha_0) e^{-\tau_0 t} + \{1 - p(\alpha_0)\} \right]^{-1}. \end{aligned} \quad (12)$$

as seen in Appendix 3.

In addition, if a truncation time is preferable, the truncated survival distribution of Exponential can be used (Gray & Tsiatis, 1989; Laska & Meisner, 1992). That is, we choose the truncated exponential distribution of form

$$\frac{e^{-\tau_0 t} - e^{-\tau_0 u_0}}{1 - e^{-\tau_0 u_0}} I(0 \leq t \leq u_0).$$

Then

$$\begin{aligned} H(t; \beta_0, \phi_0) &= \frac{\tau_0 p(\alpha_0)}{\beta_0 W_0 - \tau_0} \left(\frac{e^{-\tau_0 t} - e^{-\beta_0 W_0 t}}{1 - e^{-\tau_0 u_0}} \right) \\ &\times \left[\frac{\tau_0 p(\alpha_0)}{\beta_0 W_0 - \tau_0} \left(\frac{e^{-\tau_0 t} - e^{-\beta_0 W_0 t}}{1 - e^{-\tau_0 u_0}} \right) + p(\alpha_0) \left(\frac{e^{-\tau_0 t} - e^{-\tau_0 u_0}}{1 - e^{-\tau_0 u_0}} \right) + \{1 - p(\alpha_0)\} \right]^{-1}, \end{aligned}$$

when $0 \leq t < u_0$ and 1 otherwise. Apparently, when u_0 goes to ∞ , $H(t; \beta_0, \phi_0)$ has limit as in (12).

In addition, let $W_i(t) \equiv 1$, $\beta_0 \equiv 1$ and $\tau_0 = 1.01, 1.5, 2.0$. Fig. 1 displays the lag functions when $u_0 = \infty$. As shown in Figure 1, since there is no truncation time for the treatment effectiveness lag times, the final lag functions are smooth. When $p(\alpha_0) = 1$, i.e, 100% short-term treatment effectiveness lag times, then marginally, the treatment will eventually reaches the full effect in long run. But if there is any proportion of long-term treatment effectiveness lag times, then the full effect is not reachable, but instead, the treatment effect is washed out in long run, although it may have some effect for early period of time.

[Figure 1. about here]

Figure 2 displays the lag functions when $u_0 = 5$. Since almost all the treatment effectiveness lag times are assumed to happen before u_0 , it is not surprising to see similar patterns as in Fig. 1 before u_0 . In practice, since u_0 often serves as the termination of data collection, we should not be able to observe anything informative after u_0 . But as a demonstration, we still show the possible picture from simulation in Fig. 2 after u_0 . As shown in Fig. 2, the treatment will eventually reaches its full effect if no lag time exists.

[Figure 2. about here]

Nevertheless, as seen in both Figures 1 and 2, when τ_0 is bigger, the treatment effectiveness lag time becomes shorter, and then the mode of H tends to be reached earlier, which means the ultimate treatment effect is reached faster.

More examples of lag functions obtained from Weibull and Gamma distributions can be found in Appendix 3. In addition, the derivatives of H with respect to different parameters

are given as well.

4.2 Simulation Studies

Simulation studies have been conducted to study the performance of estimation procedure proposed in Section 3. Two covariates are generated, R , which is continuous, following the uniform distribution on $[0,1]$, and W , which is 0 or 1 with equal probability of $1/2$, mimicking a treatment indicator. The baseline hazards function is chosen to be a Weibull distribution. Lag times are generated according to mixture distribution in (1) with F_0 to be exponential and $p(\alpha_0)$ to be constant. Then Failure times are generated according to model (3), with $(\gamma_0, \beta_0) = (0, 0), (0, 1)$ and $(1, 0)$. Independent censoring times are generated from exponential distribution with different means to yield two censoring percentages of approximately 25% and 50%. Sample sizes are of 100 and 200. Estimating functions in (11) will be used for parameter estimation.

Simulation results are listed in Table 1. For each entry in the table, one thousand replicates are simulated to compute the bias and empirical coverage probability. Here, bias is defined as the difference between the sample mean of the estimates over the 1,000 simulated data sets and its respective true value; and 95% empirical coverage probability is the percentage of Wald-type 95% confidence intervals that include the true parameters. It is evident that the estimators are virtually unbiased and the nominal confidence intervals for the parameters have reasonable coverage probabilities.

[Table 1. about here]

4.3 Real Data Examples

The data to be analyzed are collected from a randomised placebo-controlled trial of the effectiveness of biodegradable carmustine polymers for treatment of recurrent brain malignant gliomas (Brem et al. 1995). After the recurrent brain tumor was removed, a medicated or placebo polymer was placed to fill in the cavity. To reach a higher local drug concentration, the medicated polymers were supposed to gradually release carmustine over a 2 to 3 week period following the placement, because it would be more effective than systematic application (Tamargo et al., 1993; Brem et al., 1995). In 27 medical centers of this trial, 222 patients were randomized to either the carmustine polymer treatment group (110 patients) or the placebo polymer group (112 patients). Their survival times measured in weeks, treatment assignment and prognostic factors can be found in Piantadosi (1997, p. 496-509). Some exploratory analysis results can be found in Brem et al. (1995) and Chen & Wang (2000).

In addition to the treatment indicator (W), another prognostic covariate of age (R) is also considered for analysis. Results from the proportional hazards model $\lambda(t|Z) = \lambda_0(t) \exp(\beta Z)$, the additive hazards model $\lambda(t|Z) = \lambda_0(t) + \beta Z$ and the proposed model (3) that assumes the exponential F_0 and logistic model for response proportions, are listed in Table 2. As shown in the table, after adjusting for age, although the treatment effect does not appear significant in either the proportional hazards model or the additive hazards model, it is significant if the treatment effectiveness lag time is taken into account. That is, given the presence of the lag time, the treatment will significantly decrease the hazard of placebo group by 0.014, adjusting for age. The estimated average treatment effectiveness lag time is about 2.502 weeks for those who have short-term lag times. Its confidence interval does not contain 0, which implies the significant presence of such lag times. In fact, as shown in Figure 1 of Chen & Wang (2000), the two groups are almost indistinguishable

till the 7th or 8th week, if the effect of age is ignored, which also graphically suggests the potential existence of such lag time. Furthermore, because of the non-significant α 's, the proportions of the short-term responders seem not varying according to treatment assignment or participant's age.

[Table 2. about here]

5 Remarks

A more general mixture model for the treatment effectiveness lag time is

$$\bar{G}(t) = p\bar{F}_1(t) + (1 - p)\bar{F}_2(t), \quad (13)$$

where $\bar{F}_1(t)$ is the survival function for p proportion of treatment effectiveness lag times and $\bar{F}_2(t)$ is the survival function for the remaining proportion, $1 - p$. To see this, let $\bar{F}_2(t) \equiv 1$, then model (13) becomes the cure mixture model. Furthermore, if the hazard function of F_1 is monotonically increasing, we should expect the hazard function of treatment effectiveness lag time to be initially increasing but decreasing later, as the p proportion of short-term treatment effectiveness lag times dropping out of the risk set with relatively more long-term treatment effectiveness lag times left.

In the potential presence of treatment effectiveness lag time, the so-called “intention-to-treat” principle (Sheiner & Rubin, 1995) may be arguable to be used to estimate the full treatment effect. For example, if $W(t)$ is a binary treatment indicator taking the value of 0 or 1 as assumed in model (3), then given the treatment effectiveness lag time U , the treatment will not be fully effective before U , i.e., it is still a true “control” before U . Therefore marginally, he or she should be equivalently counted as a member of treatment with probability of $H(t; \beta_0, \phi_0)$ exactly, although the participant is physically assigned to the

treatment group. As shown in Fig. 1, when ϕ_0 increases as in the Exponential distribution, i.e., the treatment effectiveness lag time tends to be shorter, the treatment reaches its full effect quicker.

In Model (3), the most critical part is the random effects are additive to possibly gain benefit in designing simple inference procedures. It is less critical whether or not the fixed effect of $R_i(t)$ to be additive or multiplicative. Thus another class of change point hazards models with additive random effects is:

$$\lambda\{t|Z_i(t), U_i; \theta_0\} = \lambda_0(t)e^{\gamma_0^T R_i(t)} + I(U_i \leq t)\beta_0^T W_i(t). \quad (14)$$

In contrast to the general additive-multiplicative hazards model in Lin & Ying (1995), it is not too difficult to find that model (14) is a parallel model but with random effects included in the additive component. Nevertheless, the marginalized model (14) should have same H_i 's as in (4). It is then straightforward to extend all the inference procedures and asymptotic results to model (14).

Although the model proposed in this article has certain prominent advantages, there are some critical issues in actually implementing this model. The first issue is inherited from the additive hazards model. That is, the parameter space is restricted by the magnitude of the baseline hazard function in order to obtain reasonable parameter estimates. One solution is to replace βZ with $\exp(\beta Z)$, but then the interpretation of β becomes cumbersome. The second issue is inherited from the cure mixture model, which is the potential identifiability problem with the parameters in the regression model of response proportions and the parameters in F_0 . As pointed out by Farewell (1998, p. 1051-2), the estimates of these parameters here also tend to have high correlation because of possible over-parametrisation of the lag times. This issue would be less critical if there is strong pathological evidence to support the notion of existence of two heterogeneous population. Otherwise, only modelling F_0 but ignoring p 's

is good enough to detect the potential existence of the lag time, estimate its average and derive a good lag function in practice.

APPENDIX 1

Marginalization of Model (3)

According to model (3), we know that

$$\begin{aligned}
 \Lambda(t|Z_i(t), U_i; \theta_0) &= \int_0^t \{ \lambda_0(s) + \gamma_0^\top R_i(s) + I(U_i \leq s) \beta_0^\top W_i(s) \} ds \\
 &= \Lambda_0(t) + \int_0^t \gamma_0^\top R_i(s) ds + \int_0^t I(U_i \leq s) \beta_0^\top W_i(s) ds \\
 &= \Lambda_0(t) + \int_0^t \gamma_0^\top R_i(s) ds + I(U_i \leq t) \int_{U_i}^t \beta_0^\top W_i(s) ds
 \end{aligned}$$

and

$$S(t|Z_i(t), U_i; \theta_0) = \exp \left\{ -\Lambda_0(t) - \int_0^t \gamma_0^\top R_i(s) ds - I(U_i \leq t) \int_{U_i}^t \beta_0^\top W_i(s) ds \right\}.$$

Therefore the marginal survival function is then

$$\begin{aligned}
 &S(t|Z_i(t); \theta_0, \phi_0) \\
 &= e^{-\Lambda_0(t) - \gamma_0^\top \int_0^t R_i(s) ds} \int_0^\infty e^{-\beta_0^\top I(u \leq t) \int_u^t W_i(s) ds} dG(u; \phi_0).
 \end{aligned}$$

The marginal hazard function is

$$\begin{aligned}
 \lambda(t|Z_i(t); \theta_0, \phi_0) &= -\frac{d}{dt} \log S(t|Z_i(t); \theta_0, \phi_0) \\
 &= \lambda_0(t) + \gamma_0^\top R_i(t) - \frac{d}{dt} \log \int_0^\infty e^{-\beta_0^\top I(u \leq t) \int_u^t W_i(s) ds} dG(u; \phi_0) \\
 &= \lambda_0(t) + \gamma_0^\top R_i(t) - \frac{\frac{d}{dt} \int_0^\infty e^{-\beta_0^\top I(u \leq t) \int_u^t W_i(s) ds} dG(u; \phi_0)}{\int_0^\infty e^{-\beta_0^\top I(u \leq t) \int_u^t W_i(s) ds} dG(u; \phi_0)} \\
 &= \lambda_0(t) + \gamma_0^\top R_i(t) + \frac{\beta_0^\top W_i(s) \int_0^t e^{-\beta_0^\top \int_u^t W_i(s) ds} dG(u; \phi_0)}{\int_0^\infty e^{-\beta_0^\top \int_u^t W_i(s) ds} dG(u; \phi_0) + 1 - G(t; \phi_0)}
 \end{aligned}$$

APPENDIX 2

Asymptotics

2.1 *Weak convergence of $n^{-1/2}\Gamma(t; \theta_0, \phi_0)$*

Since $n^{-1/2}\Gamma(t; \theta_0, \phi_0)$ is an \mathcal{F}_t -martingale process, with the regularity conditions of 1-3, it is true that the condition (2.5.1) and (2.5.3) of Andersen et al. (1993, p. 83) are satisfied. Hence the weak convergence of $n^{-1/2}\Gamma(t; \theta_0, \phi_0)$ is implied for any $t \in [0, t_0]$. In fact, more delicate arguments can be used as in Ying (1993) to extend t_0 to ∞ .

2.2 *Asymptotics of $n^{1/2}(\hat{\theta}^T - \theta_0^T, \hat{\phi}^T - \phi_0^T)^T$.*

It is not difficult to see that

$$\begin{aligned} & -n^{-1} \begin{pmatrix} \partial\Gamma(t_0; \theta_0, \phi_0)/\partial\theta_0 \\ \partial\Gamma(t_0; \theta_0, \phi_0)/\partial\phi_0 \end{pmatrix} \\ = & n^{-1} \sum_{i=1}^n \int_0^{t_0} Y_i(t) \{J_i(t; \theta_0, \phi_0) - \bar{J}(t; \theta_0, \phi_0)\} \{\gamma_0^T R_i(s) + \beta_0^T W_i(s) H_i(s; \theta_0, \phi_0)\}' dt \\ & - n^{-1} \sum_{i=1}^n \int_0^{t_0} \{J_i(t; \theta_0, \phi_0) - \bar{J}^*(t; \theta_0, \phi_0)\}' dP_i(t; \theta_0, \phi_0). \end{aligned}$$

The second term on the right-hand side in the above equation is an average of martingale integrals, therefore it converges in probability to zero. With the conditions in Theorem 3, it is true that the first term converges to A . Therefore, $-n\Gamma'(t_0; \theta_0, \phi_0) \xrightarrow{P} D$. In fact, following similar arguments in Lin & Ying (1995), it is also true that $(\hat{\theta}^T, \hat{\phi}^T)^T \rightarrow (\theta_0^T, \phi_0^T)^T$ consistently. Furthermore, by the Taylor expansion of $\Gamma(t_0; \hat{\theta}, \hat{\phi})$ at (θ_0, ϕ_0) , we know

$$n^{1/2} \begin{pmatrix} \hat{\theta} - \theta_0 \\ \hat{\phi} - \phi_0 \end{pmatrix} = \left\{ -n^{-1} \begin{pmatrix} \partial\Gamma(t_0; \theta^*, \phi^*)/\partial\theta^T \\ \partial\Gamma(t_0; \theta^*, \phi^*)/\partial\phi^T \end{pmatrix} \right\}^{-1} n^{-1/2}\Gamma(t_0; \theta_0, \phi_0).$$

Therefore the asymptotic normality of estimators are established. A straightforward variance calculation would lead to the results in Theorem 3.

2.3 Asymptotics of $n^{1/2}\{\hat{\Lambda}_0(t; \hat{\theta}, \hat{\phi}) - \Lambda_0(t)\}$

By the decomposition technique outlined in Fleming & Harrington (pp. 300, 1991), we have

$$\underbrace{n^{1/2}\{\hat{\Lambda}_0(t; \hat{\theta}, \hat{\phi}) - \hat{\Lambda}_0(t; \theta_0, \phi_0)\}}_{\text{I}} + \underbrace{n^{1/2}\{\hat{\Lambda}_0(t; \theta_0, \phi_0) - \hat{\Lambda}_0^*(t)\}}_{\text{II}} + \underbrace{n^{1/2}\{\Lambda_0^*(t) - \Lambda_0(t)\}}_{\text{III}}, \quad (\text{A1})$$

where $\Lambda_0^*(t) = \int_0^t I\{\sum_{i=1}^n Y_i(s) > 0\} \lambda_0(s) ds$.

Through the Taylor expansion and results in Theorem 2.2 in Lin & Ying (1995), term (I) is

$$\lim_{n \rightarrow \infty} K_1(t) D^{-1} n^{-1/2} \Gamma(t_0; \theta_0, \phi_0) + o_p(1),$$

uniformly for $t \in [0, t_0]$. In addition, term (II) is $o_p(1)$ by the Lengart inequality (Andersen, et al., 1993). It is also straightforward to see that term (III) is asymptotically ignorable. Therefore, by the multivariate martingale central limit theorem, then the asymptotic results of $\hat{\Lambda}$ is established. The variance calculation is straightforward.

APPENDIX 3

Examples of lag functions

3.1 *Exponential lag time: $f(t) = \tau_0 e^{-\tau_0 t}$*

$$\begin{aligned}
 H &= \left\{ \frac{\tau_0 p(\alpha_0)}{\beta_0 W_0 - \tau_0} (e^{-\tau_0 t} - e^{-\beta_0 W_0 t}) \right\} \left\{ \frac{\tau_0 p(\alpha_0)}{\beta_0 W_0 - \tau_0} (e^{-\tau_0 t} - e^{-\beta_0 W_0 t}) + p(\alpha_0) e^{-\tau_0 t} + 1 - p(\alpha_0) \right\}^{-1} \\
 H'_{\beta_0} &= (\tau_0 W_0 p(\alpha_0) [(1 - \tau_0 t + \beta_0 W_0 t) \{(1 - p(\alpha_0)) e^{-\beta_0 W_0 t} + p(\alpha_0) e^{-t(\tau_0 + \beta_0 W_0)}\} \\
 &\quad - (1 + p(\alpha_0) e^{-\tau_0 t} - p(\alpha_0)) e^{-\tau_0 t}]) \\
 &\quad \times \{ \tau_0 p(\alpha_0) e^{-\beta_0 W_0 t} - p(\alpha_0) e^{-\tau_0 t} \beta_0 W_0 - \beta_0 W_0 + p(\alpha_0) \beta_0 W_0 + \tau_0 - \tau_0 p(\alpha_0) \}^{-2} \\
 H'_{\tau_0} &= p(\alpha_0) [\{ (1 - p(\alpha_0)) (\tau_0^2 t - \tau_0 \beta_0 W_0 t + \beta_0 W_0) + p(\alpha_0) \beta_0 W_0 e^{-\tau_0 t} \} e^{-\tau_0 t} \\
 &\quad - (1 - p(\alpha_0)) \beta_0 W_0 e^{-\beta_0 W_0 t} + p(\alpha_0) \{ \tau_0^2 t - \tau_0 \beta_0 W_0 t - \beta_0 W_0 \} e^{-t(\tau_0 + \beta_0 W_0)}] \\
 &\quad \times (\tau_0 p(\alpha_0) e^{-\beta_0 W_0 t} - p(\alpha_0) e^{-\tau_0 t} \beta_0 W_0 - \beta_0 W_0 + \tau_0 + p(\alpha_0) \beta_0 W_0 - \tau_0 p(\alpha_0))^{-2} \\
 H'_p &= (\beta_0 W_0 - \tau_0) (e^{-\tau_0 t} - e^{-\beta_0 W_0 t}) \tau_0 \\
 &\quad \times (\tau_0 p(\alpha_0) e^{-\beta_0 W_0 t} - p(\alpha_0) e^{-\tau_0 t} \beta_0 W_0 - \beta_0 W_0 + \tau_0 + p(\alpha_0) \beta_0 W_0 - \tau_0 p(\alpha_0))^{-2}
 \end{aligned}$$

3.2 *Weibull lag time: $f(t) = \tau_{10}^{\tau_{20}} \tau_{20} t^{\tau_{20}-1} e^{-(\tau_{10} t)^{\tau_{20}}}$*

$$\begin{aligned}
 H &= p(\alpha_0) \tau_{01}^{\tau_{02}} \tau_{02} e^{-\beta_0 W_0 t} \int_0^t e^{\beta_0 W_0 u - \tau_{01}^{\tau_{02}} u^{\tau_{02}}} u^{\tau_{02}-1} du \\
 H'_{\beta_0} &= p(\alpha_0) \tau_{01}^{\tau_{02}} \tau_{02} W_0 e^{-\beta_0 W_0 t} \left\{ \int_0^t e^{\beta_0 W_0 u - (\tau_{01} u)^{\tau_{02}}} u^{\tau_{02}} du - t \int_0^t e^{\beta_0 W_0 u - (\tau_{01} u)^{\tau_{02}}} u^{\tau_{02}-1} du \right\} \\
 H'_{\tau_{01}} &= p(\alpha_0) \tau_{01}^{\tau_{02}-1} \tau_{02}^2 e^{-\beta_0 W_0 t} \left\{ \int_0^t e^{\beta_0 W_0 u - (\tau_{01} u)^{\tau_{02}}} u^{\tau_{02}-1} du - \tau_{01}^{2\tau_{02}} \int_0^t e^{\beta_0 W_0 u - (\tau_{01} u)^{\tau_{02}}} u^{2\tau_{02}-1} du \right\} \\
 H'_{\tau_{02}} &= p(\alpha_0) \tau_{01}^{\tau_{02}} e^{-\beta_0 W_0 t} \left[\{1 + (\log \tau_{01}) \tau_{02}\} \int_0^t e^{\beta_0 W_0 u - (\tau_{01} u)^{\tau_{02}}} u^{\tau_{02}-1} du \right. \\
 &\quad \left. + \tau_{02} \int_0^t e^{\beta_0 W_0 u - (\tau_{01} u)^{\tau_{02}}} (-\tau_{01}^{\tau_{02}} \log \tau_{01} u^{\tau_{02}} - \tau_{01}^{\tau_{02}} u^{\tau_{02}} \log u + \log u) u^{\tau_{02}-1} du \right] \\
 H'_p &= \tau_{01}^{\tau_{02}} \tau_{02} e^{-\beta_0 W_0 t} \int_0^t e^{\beta_0 u W_0 - (\tau_{01} u)^{\tau_{02}}} u^{\tau_{02}-1} du
 \end{aligned}$$

3.3 Gamma lag time: $f(t) = \Gamma^{-1}(\tau_{02}) \tau_{01}^{\tau_{02}} t^{\tau_{02}-1} e^{-\tau_{01} t}$

$$\begin{aligned}
H &= \frac{\tau_{01}^{\tau_{02}} p(\alpha_0) e^{-\beta_0 W_0 t}}{\Gamma(\tau_{02})} \int_0^t u^{\tau_{02}-1} e^{(\beta_0 W_0 - \tau_{01})u} u^{\tau_{02}-1} du \\
H'_{\beta_0} &= \frac{-\tau_{01}^{\tau_{02}} p(\alpha_0) W_0 e^{-\beta_0 W_0 t}}{\Gamma(\tau_{02})} \left[t \int_0^t e^{(\beta_0 W_0 - \tau_{01})u} u^{2(\tau_{02}-1)} du - \int_0^t e^{(\beta_0 W_0 - \tau_{01})u} u^{2\tau_{02}-1} du \right] \\
H'_{\tau_{01}} &= \frac{\tau_{01}^{\tau_{02}-1} p(\alpha_0) e^{-\beta_0 W_0 t}}{\Gamma(\tau_{02})} \left[\tau_{02} \int_0^t e^{(\beta_0 W_0 - \tau_{01})u} u^{2(\tau_{02}-1)} du - \tau_{01} \int_0^t e^{(\beta_0 W_0 - \tau_{01})u} u^{2\tau_{02}-1} du \right] \\
H'_{\tau_{02}} &= -\frac{\Gamma'(\tau_{02}) \tau_{01}^{\tau_{02}} p(\alpha_0) e^{-\beta_0 W_0 t}}{\Gamma^2(\tau_{02})} \int_0^t u^{\tau_{02}-1} e^{(\beta_0 W_0 - \tau_{01})u} u^{\tau_{02}-1} du \\
&\quad + \frac{\tau_{01}^{\tau_{02}} p(\alpha_0) e^{-\beta_0 W_0 t}}{\Gamma(\tau_{02})} \left[\log \tau_{01} \int_0^t e^{(\beta_0 W_0 - \tau_{01})u} u^{2\tau_{02}-2} du + 2 \int_0^t e^{(\beta_0 W_0 - \tau_{01})u} u^{2\tau_{02}-2} \log u du \right] \\
H'_p &= \frac{\tau_{01}^{\tau_{02}} e^{-\beta_0 W_0 t}}{\Gamma(\tau_{02})} \int_0^t e^{(\beta_0 W_0 - \tau_{01})u} u^{2\tau_{02}-2} du
\end{aligned}$$

References

- AALLEN, O. O. (1980). A model for nonparametric regression analysis of counting processes. In *Lecture Notes in Statistics* **2**, New York: Springer-Verlag.
- BASU, A. P., GHOSH, J. K. & JOSHI, S. N. (1988). On estimating change point in a failure rate. *Statist. Decis.* **4**, 239-52.
- BREM, H., PIANTADOSI, S., BURGER, P. C., WALKER, M., SELKER, R., VICK, N.A., BLACK, K., SISITI, M., BREM, S., MOHR, G., MULLER, P. & MORAWETZ, R. (1995). Placebo-controlled trials of safety and efficacy of intraoperative controlled delivery by biodegradable polymers of chemotherapy of recurrent gliomas. *The Lancet* **345**, 1008-12.
- BRESLOW, N.E., DAY, N.E. (1980). *Statistical Methods in Cancer Research II: The Design and Analysis of Case-Control Studies*. Lyon: IARC.
- BRESLOW, N.E., DAY, N.E. (1987). *Statistical Methods in Cancer Research I: The Design and Analysis of Cohort Studies*. Lyon: IARC.
- CHEN, Y.Q. & WANG, M.-C. (2000). Analysis of accelerated hazards model. *J. Amer. Statist. Assoc.* **95**, 608-18.
- COX, D. R. (1972). Regression models and life tables (with Discussion). *J. R. Statist. Soc. B* **34**, 187-220.
- GRAY, R. J. & TSIATIS, A. A. (1989). A linear ran test for use when the main interest is in differences in cure rates. *Biometrics* **45**, 899-904.
- FAREWELL, V. T. (1982). The use of mixture models for the analysis of survival data with long-term survivors. *Biometrics* textbf38, 1041-6.

- FAREWELL, V. T. (1998). Cure Models. In *Encyclopedia of Biostatistics* (Ed. Peter Armitage and Theodore Colton, p. 1051-3). Chichester: John Wiley.
- FLEMING, T.R. & HARRINGTON, D.P. (1991). *Counting Processes and Survival Analysis*. New York: John Wiley.
- GAIL, M.H. (1985). Applicability of sample size calculations based on a comparison of proportions for use with the Log Rank test. *Contr. Clin. Trials* **6**, 112-9.
- LAKATOS, E. (1986). Sample size determination in clinical trials with time-dependent rates of losses and noncompliance. *Contr. Clin. Trials* **7**, 189-99.
- LAI, T.L. & YING, Z. (1992). Asymptotically efficient estimation in censored and truncated regression models. *Statist. Sinica* **2**, 17-46.
- LASKA, E. M. & MEISNER, M. J. (1992). Nonparametric estimation and testing in a cure model. *Biometrics* **48**, 1223-34.
- LIN, D.Y. & YING, Z. (1994). Semiparametric analysis of the additive risk model. *Biometrika* **81**, 61-71.
- LIN, D.Y. & YING, Z. (1995). Semiparametric analysis of general additive-multiplicative hazard models for counting processes. *Ann. Stat.* **5**, 1712-34.
- LIN, D.Y. & YING, Z. (1997). Additive hazards regression models for survival data. In *Lecture Notes in Statistics* **123** (Ed. D. Y. Lin and T. R. Fleming, p. 185-198). New York: Springer-Verlag.
- MCCULLAGH, P. & NELDER, J.A. (1989). *Generalized Linear Models*, 2nd ed. New York: Chapman and Hall.

- MCKEAGUE, I.W. (1992). Discussion of Paper by P.D. Sasieni. *Survival Analysis: State of the Art*. Dordrecht: Kluwer Academic Publishers.
- NGUYEN, H. T., ROGERS, G. S. & WALKER, E. A. (1984). Estimation in change-point hazard rate models. *Biometrika* **71**, 299-304.
- PÉREZ, V., GILABERTE, I., FARIES, D., ALVAREZ, E, & ARTIGAS, F (1997). Randomised, double-blind, placebo-controlled trial of pindolol in combination with fluoxetine antidepressant treatment. *The Lancet* **349**, 1594-97.
- PIANTADOSI, S. (1997). *Clinical Trials: A Methodologic Perspective*. New York: John Wiley.
- SELF, S., PRENTICE, R., IVERSON, D., HENDERSON, M., THOMPSON, D., BYAR, D., INSULL, W., GORBACH, S. I., CLIFFORD, C., GOLDMAN, S., URBAN, N., SHEPPARD, L. & GREENWALD, P. (1988). Statistical design of the women's health trial. *Contr. Clin. Trials* **9**, 119-36.
- SHEINER, L.B. & RUBIN, D.B. (1995). Intention-to-treat analysis and the goals of clinical trials, *Clinical Pharmacology and Therapeutics* **57**, 6-15.
- TAMARGO, R.J., MYSEOS, J.S., EPSTEIN, J.I., YANG, M.B., CHASIN, M. & BREM, H. (1993). Interstitial chemotherapy of the 9L gliosarcoma: controlled release polymers for drug delivery in the brain. *Cancer Res* **53**:329-33.
- TAMURA, R.N., FARIES, D.E. & FENG, J. (2000). Comparing time to onset of response in antidepressant clinical trials using the cure model and the Cramer-von Mises test. *Statist. Med.* **19**, 2169-84.
- WU, M., FISHER, M., & DEMETS, D. (1980). Sample sizes for long-term medical trial with time-dependent dropout and event rates. *Contr. Clin. Trials* **1**, 109-21.

ZUCKER, D.M. & LAKATOS, E. (1990). Weighted Log Rank type statistics for comparing survival curves when there is a time lag in the effectiveness of treatment. *Biometrika* **77**, 853-64.

Table 1: Summary of simulation studies. Each entry is the estimated bias, with the associated 95% empirical coverage probability in brackets.

n	%	$(\gamma_0, \beta_0) = (0, 0)$		$(\gamma_0, \beta_0) = (0, 1)$		$(\gamma_0, \beta_0) = (1, 0)$	
		R	W	R	W	R	W
100	25%	-0.056(0.950)	-0.038(0.962)	-0.005(0.959)	0.042(0.941)	-0.056(0.948)	0.024(0.955)
	50%	0.078(0.927)	0.021(0.955)	-0.058(0.941)	0.026(0.930)	-0.059(0.939)	0.047(0.974)
200	25%	0.008(0.949)	-0.029(0.944)	-0.006(0.961)	-0.021(0.960)	-0.025(0.955)	0.004(0.941)
	50%	-0.026(0.951)	-0.008(0.941)	-0.025(0.946)	-0.015(0.964)	0.014(0.945)	0.020(0.935)

%, censoring percentage

Table 2: Regression analysis of brain cancer trial accounting for treatment effectiveness lag time

Model	Parm	Cov	PHM			ADHM			ADHM-L		
			Est	SE	95%CI	Est	SE	95%CI	Est	SE	95%CI
Hazard λ	β	Treatment	-0.227	0.140	(-0.501,0.047)	-0.005	0.003	(-0.001,0.011)	-0.014	0.009	(-0.032,-0.028)
	γ	Age	0.221	0.054	(0.115,0.327)	0.005	0.001	(0.003,0.007)	0.006	0.001	(0.004,0.008)
Response p	α	Treatment							-0.002	3.063	(-6.005,6.001)
		Age							-0.277	0.509	(-1.275,0.721)
F_0	μ							2.502	1.037	(0.469,4.535)	

PHM, proportional hazards model; ADHM, additive hazards model; ADHM-L, additive hazards model with latent treatment effectiveness lag time; Parm, parameter; Cov, covariate; Est, estimate; SE, estimated standard error; 95%CI, 95% confidence interval.

Figure 1: Lag function of $H(t; \beta_0, \phi_0)$ without truncation time

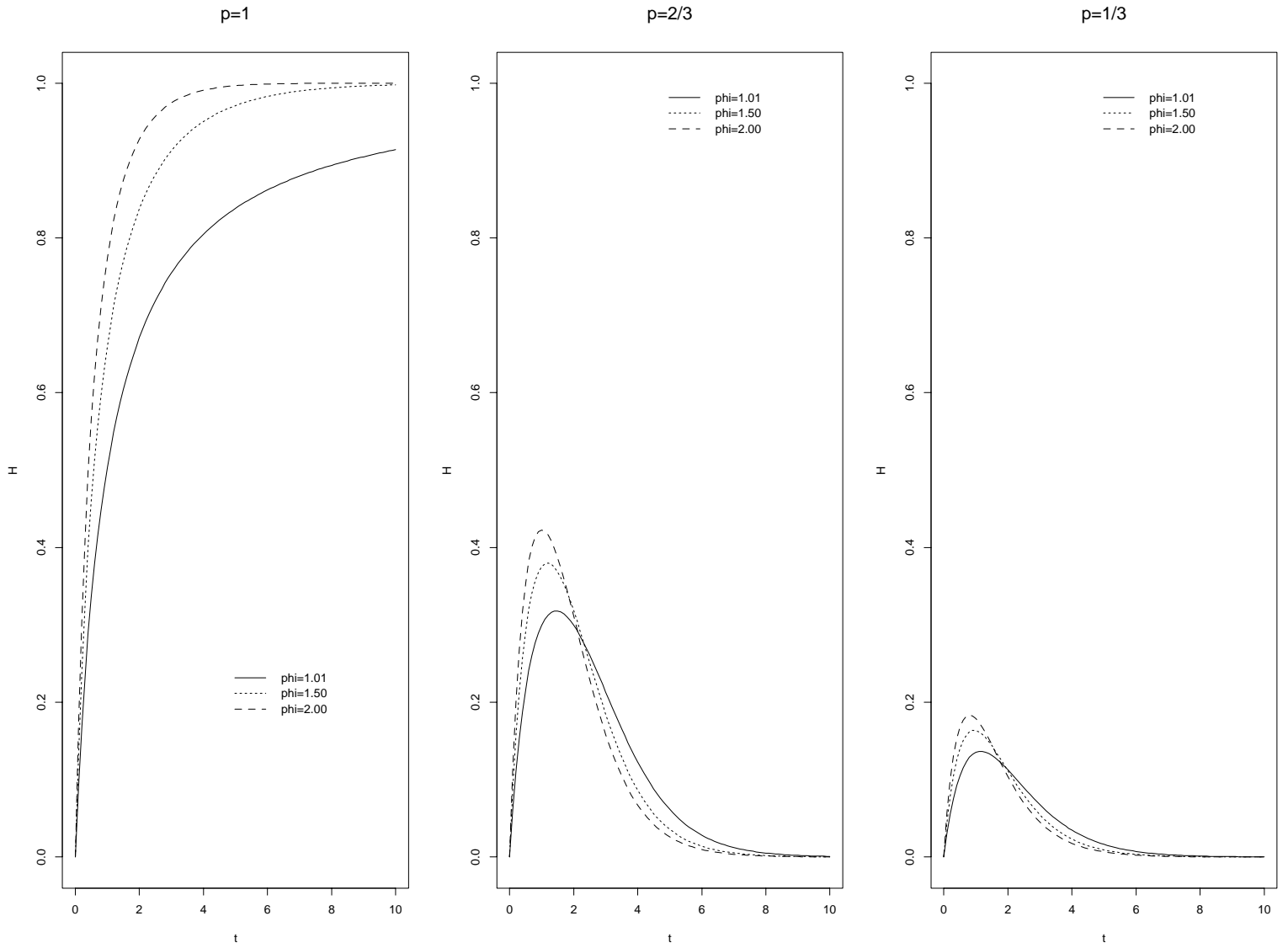


Figure 2: Lag function of $H(t; \beta_0, \phi_0)$ with truncation time

