Inference for proportional mean residual life model in presence of censoring

By Y. Q. CHEN and N. P. JEWELL

Division of Biostatistics, School of Public Health, University of California

Berkeley, California 94720, U.S.A.

yqchen@stat.berkeley.edu jewell@stat.berkeley.edu

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

Inference for proportional mean residual life model in presence of censoring

Summary

As a function of time t, mean residual life is defined as remaining life expectancy of a subject given its survival to t. It plays an important role in many research areas to characterise stochastic behavior of survival over time. Similar to the Cox proportional hazard model, the proportional mean residual life model were proposed in statistical literature to study association between the mean residual life and individual subject's explanatory covariates. In this article, we will study this model and develop appropriate inference procedures in presence of censoring. Numerical studies including simulation and real data analysis are presented as well.

Some key words: Censored survival data; Counting process; Equilibrium renewal process; Semiparametric regression model.

1 Introduction

Time-to-event data or survival data have been collected and studied in many research areas for decades. Although the full length of well-defined survival time is always of interest to study, the residual life time of a subject at time t, which is the remaining survival time given the subject surviving up to t, is also important to study in areas such as industrial reliability, demography and life insurance.

In statistical literature, the most important function regarding to the residual life is mean residual life:

$$m(t) = E(T - t|T > t)$$

for $t \geq 0$, where T is the survival time, a nonnegative real-value random variable (Chiang, 1960). That is, the mean residual life is the remaining life expectancy given survival to t. For a comprehensive review of previous research on this function, readers are refereed to Guess & Proschan (1988) and Csörgő & Zitikis (1996).

To evaluate the association between the mean residual life and its covariates Z, the proportional mean residual life model is proposed (Oakes & Dasu, 1990; Zahedi, 1991):

$$m(t|Z) = m_0(t)\exp(\beta^{\mathrm{T}} Z), \tag{1}$$

where $m(\cdot|Z)$ is the mean residual life of $p-$vector covariate Z, $m_0(t)$ is some unknown baseline mean residual life of covariate $Z = 0$, and $\beta \in \mathcal{B} \subset \mathbf{R}^p$ is parameter. Here the superscript T denotes vector transpose.

Model (1) is closely related to the accelerated failure time model (Kalbfleisch & Prentice,

1980):

$$\log T = \beta^{\mathrm{T}} Z + \epsilon, \tag{2}$$

where ϵ's are i.i.d. random errors with an unspecified distribution. To see this, let $t = 0$ in model (1), then

$$\log E(T|Z) = \log m(t = 0|Z) = \log m_0(0) + \beta^{\mathrm{T}} Z.$$

Although $E(\log T|Z)$ and $\log E(T|Z)$ are not necessarily identical by the Jensen's Inequality, careful choice of the baseline mean residual life in (1) and the distribution of ϵ in (2) would lead to identical models. Given the conditions for validity of $m_0(t)$ in Hall & Wellner (1981), the proof in Appendix 1 shows that

PROPOSITION 1. *Suppose that there is a sequence of constants* $\{d_k\}_{k=-\infty}^{+\infty}$ *such that* $m_0(t) = \sum_{k=-\infty}^{+\infty} d_k t^k$ *on* $[0, t_0]$ *for some big enough* $t_0 > 0$. *Then the proportional mean residual life model of (1) and the accelerated failure time model of (2) coincide if and only if* $m_0(t)$ *is constant.*

Model (1) is also related to the proportional hazards model (Cox, 1972). As noticed in Meilijson (1972), the hazard function of the forward recurrence time, V, say, in the equilibrium renewal processes formed by T's is the reciprocal of the mean residual life of T (Cox, 1962, p. 27) . Thus,

$$\lambda_v(t|Z) = \lambda_{v,0}(t) \exp(-\beta^{\mathrm{T}} Z), \tag{3}$$

where $\lambda_v(\cdot)$ stands for hazard function of the forward recurrence time at a fixed time for equilibrium renewal process that has the same underlying distribution as T. As shown in Oakes & Dasu (1990, Theorem 2) and Gupta & Kirmani (1998, Theorems 2.1 & 2.2), the sufficient and necessary condition for which the proportional mean residual life model and the proportional hazards model coincide is that when $m_0(t)$ is linear in t. In conjunction

with Proposition 1, it is true that the afore-mentioned three models are identical if and only if the underlying distribution of T is exponential.

Although inference on the parameters in model (1) was briefly discussed in Zahedi (1991), it is mainly maximum likelihood method for complete survival data when the baseline mean residual life is parametric. Elegant approaches by Maguluri and Zhang (1994) utilized the relationship in (3) when the baseline mean residual life is unknown, but they are restricted to the situation when survival times are not censored. In practice, however, survival data collected for analysis often involves censoring, when the complete survival time is not fully observable. In this article, we will propose appropriate inference procedures in presence of censoring. Details are elaborated in §2. Numerical studies including Monte-Carlo simulation and actual data analysis are in §3. Some remarks are in §4. Technical proofs are collected in Appendices.

2 Inference Procedures in Presence of Censoring

Let C be the potential censoring time. Conditional on Z, T and C are assumed to be independent. Suppose that the observed data set consist of n independent triplets of (X_i, Δ_i, Z_i), where $X_i = \min(T_i, C_i)$, $\Delta_i = I(T_i \leq C_i)$ for $i = 1, 2, \ldots, n$. Here , $I(\cdot)$ is the indicator function taking the value of 1 if the condition is satisfied and 0 otherwise. In addition, C_i are assumed to be homogeneous and follow the survival function of $F_c(\cdot)$.

When there is no censoring, the approach in §3.2 of Maguluri and Zhang (1994) provides a way to estimate the parameters in model (1). Specifically, they first noticed that the

forward recurrence time V of covariate Z has survival function:

$$S_v(t|Z) = E\{I(V \geq t)|Z\} = \frac{E\{(T-t)^+|Z\}}{m_0(t)\exp(\beta^{\mathrm{T}} Z)},$$

where $(T-t)^+$ denotes $(T-t)I(T \geq t)$. And for complete V, the partial score estimating equation based on the proportional hazards model (3) is:

$$\hat{E}(Z) - \int_0^\infty \frac{\hat{E}\{Z\exp(-\beta^{\mathrm{T}} Z)I(V \geq t)\}}{\hat{E}\{\exp(-\beta^{\mathrm{T}} Z)I(V \geq t)\}} d\hat{F}_v(t) = 0, \tag{4}$$

where $\hat{F}_v(\cdot)$ is an appropriate estimator of the distribution function $F(\cdot)$ of V. Therefore, with replacement of $I(V \geq t)$ with an appropriate estimator, an estimating equation of β based on T is

$$\hat{E}(Z) - \int_0^\infty \frac{\hat{E}\{Z\exp(-2\beta^{\mathrm{T}} Z)(T-t)^+\}}{\hat{E}\{\exp(-2\beta^{\mathrm{T}} Z)(T-t)^+\}} \cdot \frac{\hat{E}\{\exp(-\beta^{\mathrm{T}} Z)I(T > t)\}}{\hat{E}\{\exp(-\beta^{\mathrm{T}} Z)T\}} dt. \tag{5}$$

Therefore, by replacing $\hat{E}$ with its respective empirical estimator, an estimating function was proposed as

$$n^{-1}\sum_{i=1}^n Z_i - \int_0^\infty \frac{\sum_i Z_i \exp(-2\beta^{\mathrm{T}} Z_i)(T_i - t)^+}{\sum_i \exp(-2\beta^{\mathrm{T}} Z_i)(T_i - t)^+} \cdot \frac{\sum_i \exp(-\beta^{\mathrm{T}} Z_i)I(T_i > t)}{\sum_i \exp(-\beta^{\mathrm{T}} Z_i)T_i} dt. \tag{6}$$

Asymptotic results regarding to estimator $\hat{\beta}$ in (6) were derived as well.

However, when censoring presents, T_i are not always observable but potentially censored. That is, (X_i, Δ_i) are often observed, instead. Then equation (6) needs to be modified to accommodate such potential censoring. It is not difficult to see that, for any well-defined function of $H(X_i, Z_i, t)$,

$$\begin{aligned} E\left\{\frac{H(X_i, Z_i, t)\Delta_i}{F_c(X_i)}\right\} &= E\left[E\left\{\frac{H(X_i, Z_i, t)\Delta_i}{F_c(X_i)}\middle| Z_i\right\}\right] \\ = \quad E\left(E\left[\frac{H(T_i, Z_i, t)}{F_c(T_i)} E\{I(C_i \geq T_i)\}\middle| Z_i, T_i\right]\right) &= E\left\{H(T_i, Z_i, t)\right\}. \end{aligned}$$

Therefore, if let $H(X_i, Z_i, t) \equiv Z_i$, $H(X_i, t) = (X_i - t)^+$ and $H(X_i, t) = I(X_i \geq t)$, we will have

$$E\left\{\frac{Z_i \Delta_i}{F_c(X_i)}\right\} = E(Z_i),$$

$$E\left\{\frac{(X_i - t)^+ \Delta_i}{F_c(X_i)}\right\} = E\left\{(T_i - t)^+\right\}$$

and

$$E\left\{\frac{I(X_i \geq t)\Delta_i}{F_c(X_i)}\right\} = E\left\{I(T_i \geq t)\right\},$$

respectively. So, it is natural to further replace Z_i, $(T_i - t)^+$ and $I(T_i \geq t)$ with $Z_i \Delta_i \{\hat{F}_c(X_i)\}^{-1}$, $(X_i - t)^+ \Delta_i \{\hat{F}_c(X_i)\}^{-1}$ and $I(X_i \geq t)\Delta_i \{\hat{F}_c(X_i)\}^{-1}$ in equation (6), respectively, where $\hat{F}_c(\cdot)$ is an appropriate estimator of $F_c(\cdot)$, such as its Kaplan-Meier estimator. Then the resulting estimating function denoted as $U(\beta)$ is

$$\begin{aligned} U(\beta) &= n^{-1} \sum_{i=1}^{n} \frac{Z_i \Delta_i}{\hat{F}_c(X_i)} \\ - & \int_0^\infty \frac{\sum_i Z_i \exp(-2\beta^{\mathrm{T}} Z_i)(X_i - t)^+ \Delta_i \{\hat{F}_c(X_i)\}^{-1}}{\sum_i \exp(-2\beta^{\mathrm{T}} Z_i)(X_i - t)^+ \Delta_i \{\hat{F}_c(X_i)\}^{-1}} \cdot \frac{\sum_i \exp(-\beta^{\mathrm{T}} Z_i) I(X_i > t) \Delta_i \{\hat{F}_c(X_i)\}^{-1}}{\sum_i \exp(-\beta^{\mathrm{T}} Z_i) X_i \Delta_i \{\hat{F}_c(X_i)\}^{-1}} dt. \\ = & \; n^{-1} \sum_{i=1}^{n} \frac{\Delta_i}{\hat{F}_c(X_i)} \left\{ Z_i - \bar{Z}_i(\beta, \hat{F}_c) \right\} \end{aligned}$$

where

$$\bar{Z}_i(\beta, \hat{F}_c) = \int_0^\infty \frac{\sum_i Z_i \exp(-2\beta^{\mathrm{T}} Z_i)(X_i - t)^+ \Delta_i \{\hat{F}_c(X_i)\}^{-1}}{\sum_i \exp(-2\beta^{\mathrm{T}} Z_i)(X_i - t)^+ \Delta_i \{\hat{F}_c(X_i)\}^{-1}} \cdot \frac{\exp(-\beta^{\mathrm{T}} Z_i) I(X_i > t) dt}{n^{-1} \sum_i \exp(-\beta^{\mathrm{T}} Z_i) X_i \Delta_i \{\hat{F}_c(X_i)\}^{-1}}.$$

Denote $\hat{\beta}$ the solution to $U(\beta) = 0$.

Using a martingale representation of the Kaplan-Meier estimator $\hat{F}_c$ (Fleming & Harrington, 1991, p. 97), some straightforward algebraic manipulation with the functional

Delta-method shows that

$$\begin{aligned}\frac{\Delta_i}{\hat{F}_c(X_i)} &\simeq \frac{\Delta_i}{F_c(X_i)} + \frac{\Delta_i}{F_c(X_i)} \cdot \frac{F_c(X_i) - \hat{F}_c(X_i)}{F_c(X_i)} \\ &\simeq \frac{\Delta_i}{F_c(X_i)} + n^{-1} \sum_{j=1}^{n} \int_0^\infty \frac{\Delta_i I(X_i \geq t)}{F_c(X_i)} \frac{\hat{F}_c(t-)}{F_c(t)} \frac{dM_j(t)}{B_n(t)},\end{aligned}$$

where

$$M_j(t) = I(X_j \leq t, \Delta_j = 0) - \int_0^t I(X_j \geq s) d\Lambda_c(s),$$

and

$$B_n(t) = n^{-1} \sum_{i=1}^{n} I(X_i \geq t) \to B(t).$$

Here $\Lambda_c = -\log(1 - F_c)$ is the cumulative hazard function of the censoring times. Denote $\mathcal{F}_c(t)$ the filtration generated by the σ-algebras

$$\sigma\{I(C_i \leq s), s \leq t; I(T_i \leq u), Z_i(u), 0 \leq u < \infty, i = 1, 2, \ldots, n\}.$$

Then $\{M_j(t), j = 1, 2, \ldots, n\}$ are martingales with respect to the filtration $\mathcal{F}_c$. Therefore, according to the Law of Large Numbers, it is also true that $U(\beta) \to u(\beta)$, where

$$\begin{aligned}u(\beta) &= E\left[\frac{\Delta}{F_c(X)}\{Z - \bar{Z}(\beta, F_c)\}\right] \\ &= E(Z) - \int_0^\infty \frac{E\{Z \exp(-2\beta^{\mathrm{T}} Z)(T-t)^+\}}{E\{\exp(-2\beta^{\mathrm{T}} Z)(T-t)^+\}} \cdot \frac{E\{\exp(-\beta^{\mathrm{T}} Z) I(T > t)\}}{E\{\exp(-\beta^{\mathrm{T}} Z) T\}} dt,\end{aligned}$$

and $u(\beta_0) = 0$. Here β_0 is the true value of parameter β. According to Proposition 1 in Maguluri and Zhang (1994), it is known that $u'(\beta_0)$ is negative definite. Therefore, $u(\beta_0) = 0$ implies that $\hat{\beta}$ is consistent and unique in the neighbourhood of β_0.

To obtain a large-sample approximation of the distribution of $\hat{\beta}$, a Taylor expansion shows that $n^{1/2}(\hat{\beta} - \beta_0)$ is asymptotically equivalent to $\partial U(\beta_0)/\partial\beta \cdot n^{1/2} U(\beta_0)$. Furthermore, assume that $\mathcal{B}$ is a closed set of β. Then $U(\beta) \to u(\beta)$ uniformly and hence $\partial U(\beta_0)/\partial\beta \to u'(\beta_0)$.

With the asymptotic normality of $n^{1/2}U(\beta_0)$ established in Appendix 2, we have the following result

$$n^{1/2}(\hat{\beta}-\beta_0) \xrightarrow{\mathcal{D}} \mathrm{N}(0, \{u'(\beta_0)\}^{-1}\,\Sigma(\beta_0, F_c)\left[\{u'(\beta_0)\}^{-1}\right]^{\mathrm{T}}).$$

Therefore, the asymptotic distribution of $\hat{\beta}$ can be approximated by a normal with mean β_0 and variance-covariance matrix estimated by replacing theoretical terms in the above variance-covariance with their empirical estimates, that is,

$$\left\{\frac{\partial U(\hat{\beta})}{\partial \beta}\right\}^{-1} \hat{\Sigma}(\hat{\beta}, \hat{F}_c)\left[\left\{\frac{\partial U(\hat{\beta})}{\partial \beta}\right\}^{-1}\right]^{\mathrm{T}}.$$

More generally, a weighted version of $U(\beta)$ which includes weight function can be proposed as:

$$U^w(\beta) = n^{-1}\sum_{i=1}^{n} \frac{w(\beta^{\mathrm{T}}Z_i)\Delta_i}{\hat{F}_c(X_i)}\{Z_i - \bar{Z}_i(\beta, \hat{F}_c)\}, \tag{7}$$

where $w(\cdot)$ is a specified weight function. Appropriate weight function can be selected in $U^w(\beta)$ to allow $\partial U^w(\beta)/\partial\beta$ to be negative definite for $\beta \in \mathcal{B}$ (Maguluri and Zhang 1994), then its monotonicity and hence the uniqueness of $\hat{\beta}$ of the estimating function can be obtained as well. Another application of the weighted estimating function of $U^w(\beta)$ is to assess adequacy of the proportional mean life model in (1). Similar to the methods proposed in Gill & Schumacher (1987) and Lin (1991) to assess the adequacy of the proportional hazards model, we can choose two different weight functions, $w_1(\beta^{\mathrm{T}}Z_i)$ and $w_2(\beta^{\mathrm{T}}Z_i)$, say. Then we can test the null hypothesis of adequacy of the proportional hazards model through the quadratic form test statistics

$$T_{\mathrm{GSL}} = (\hat{\beta}_1^w - \hat{\beta}_2^w)^{\mathrm{T}}\hat{\Sigma}_{12}^{-1}(\hat{\beta}_1^w - \hat{\beta}_2^w),$$

where $\hat{\beta}_1^w$ and $\hat{\beta}_2^w$ are corresponding solutions respectively, and $\hat{\Sigma}_{12}$ is an appropriate estimator of the variance-covariance matrix of $\hat{\beta}_1^w - \hat{\beta}_2^w$. In practice, a simulation approach by Lin

(1998) can be adapted to obtain $\hat{V}$ in practice. The proposed statistics is asymptotically χ^2_p under the null hypothesis.

One limitation with the proposed inference procedure is that the censoring times follows same marginal distribution. Although this usually may not be a serious issue in controlled randomised clinical trials, it may be challenged in practice when the distributions of censoring times are different among individuals. One remedy is to sort the covariates into groups and estimate survival function for each individual group. Then it is straightforward to extend the proposed inference procedure to multiple groups. Specifically, a grouping version of estimating function $U(\beta)$, $U_g(\beta)$ is proposed as

$$U_g(\beta) = n^{-1} \sum_{i=1}^{n} \frac{\Delta_i}{\hat{F}_{c,i}(X_i)} \{Z_i - \bar{Z}_i(\beta, \hat{F}_{c,i})\}$$

where $\hat{F}_{c,i}$ is the Kaplan-Meier estimator of the censoring times of the covariate group that Z_i belongs to, and

$$\begin{aligned} \bar{Z}_i(\beta, \hat{F}_{c,i}) &= \int_0^\infty \frac{\sum_i Z_i \exp(-2\beta^{\mathrm{T}} Z_i)(X_i - t)^+ \Delta_i \{\hat{F}_{c,i}(X_i)\}^{-1}}{\sum_i \exp(-2\beta^{\mathrm{T}} Z_i)(X_i - t)^+ \Delta_i \{\hat{F}_{c,i}(X_i)\}^{-1}} \\ &\quad \times \frac{\exp(-\beta^{\mathrm{T}} Z_i) I(X_i > t) dt}{n^{-1} \sum_i \exp(-\beta^{\mathrm{T}} Z_i) X_i \Delta_i \{\hat{F}_{c,i}(X_i)\}^{-1}}. \end{aligned}$$

Furthermore, the asymptotic properties of $\hat{\beta}_g$ solved in $U_g(\beta) = 0$ follow similar argument as used for $\hat{\beta}$. However, when the covariates are continuous or of high dimension, it is almost impossible to estimate individual survival curves nonparametrically because of the so-called "curse of dimensionality." More sophisticated approach involving modeling the censoring times through a proportional hazards model can then be adapted (van der Laan, Gill & Robins, 2000). Brief discussion of this approach is laid out in §4.

3 Numerical Studies

Several simulation studies are conducted to evaluate the proposed inference procedures. To demonstrate our inference procedures in presence of censoring, we adopt similar simulation setup as in Maguluri and Zhang (1994). We mimic two-arm randomised clinical trials with approximately $n/2$ ($n = 100, 200$) subjects being randomised into each arm. That is, covariate Z is simulated to be a binary indicator with success probability of 0.5. The choices of parameter β_0 are -1, 0 and 1. The baseline distribution has linear mean residual life of the Hall-Wellner Family:

$$m_0(t) = (D_1 t + D_2)^+,$$

where $D_1 > -1$ and $D_2 > 0$. We fix $D_2 = 1$, and let $D_1 = -0.5, 0$ and 0.5, corresponding to a rescaled Beta, a unit exponential and a Pareto distribution, respectively (Oakes & Dasu, 1990). In addition, independent censoring times are generated from exponential distribution with different mean μ's which allow different prespecified censoring proportions: 0%, 10%, 20% and 30%.

Summary of simulation results is in Table 1. The quantity in each entry is computed based on 500 simulations of its corresponding configuration. In Table 1, bias is the absolute difference between the sample mean of the estimates of 500 simulations and the true value, and coverage probability is the percentage of the Wald-type 95% confidence intervals that contain the true parameter. As shown in Table 1, $\hat{\beta}$ is virtually unbiased and carries reasonable coverage probabilities.

[Table 1 about here]

A survival data set collected in actual randomised clinical trial is used as an example. The trial was conducted by Lad, et al. (1988) to evaluate the efficacy of systematic combination of chemotherapy for incompletely resected non-small-cell lung cancer. In this trial, a total of 172 participants were randomised to receive either postoperative radiotherapy (RT) solely or postoperative RT and chemotherapy with Cytoxan, Adriamycin and Platinol (CAP) for six months. One of the endpoints to evaluate the efficacy is survival time since randomisation. A log-rank test statistic to compare the survivals of two treatment arms is 9.80 with p-value of 0.002.

Furthermore, let Z be the group indicator of postoperative RT and chemotherapy with Cytoxan, Adriamycin and Platinol (CAP). If the proportional hazards model

$$\lambda(t|Z=1) = \lambda(t|Z=0)\exp(\beta)$$

is fitted, the partial likelihood estimate of β is -0.57 with estimated standard error of 0.18 (hazard ratio of 0.56 with p-value$<$0.05). Therefore, the combined treatment is seemly associated with lower hazard of survival time and hence may lead to longer survival time. However, the proportionality of hazard functions of two treatment arms may not be satisfied for the proportional hazards model, as shown in Figure 1 of the smoothed hazard functions (smoothing bandwidth is 90 days).

[Figure 1 about here]

Alternatively, a proportional mean residual life model

$$m(t|Z=1) = m(t|Z=0)\exp(\beta) \tag{8}$$

is to be fitted. It is known that the censoring proportions in the treatment and control groups are 33.3% and 18.6%, respectively. Furthermore, the log-rank statistic is 0.274 with

p-value of 0.601. And, a graphical display of Kaplan-Meier estimates of censoring survival functions does not show strong pattern of differential. Therefore, it is sensible to use $U(\beta)$ to estimate parameter β. In fact, $\hat{\beta} = 0.61$ with standard error of 0.17 (p-value< 0.05). This also suggests a significant benefit of the treatment in prolonging the mean residual life of lung cancer patients, for which the treatment mean residual life is 1.84 times of the control's.

[Figure 2 about here]

In addition to the treatment indicator, three more covariates are selected into model (8): Cell Type (squamous versus non-squamous), Age and Gender. The estimates of additional regression parameters are -0.47 ($s.e. = 0.20$, p-value< 0.05), -0.002 ($s.e. = 0.01$, p-value$>$ 0.5) and 0.12 ($s.e. = 0.20$, p-value> 0.5), while the parameter estimate of treatment indicator is 0.65 ($s.e. = 0.17$, p-value< 0.05). This is consistent with what was found with the proportional hazards model in Piantadosi (1997, p. 310). That is, by adjusting for potential confounding variables of age and gender, the treatment indicator and the cell type stand out as significant variables. The adjusted treatment effect still shows benefit of the combined treatment in prolonging mean residual life.

4 Remarks

In contrast to the widely used proportional hazards model (Cox, 1972), the proportional mean residual life model is more intuitively appealing to practitioners because of its interpretation in conditional mean. However, similar to assumption of the proportional hazards model, the proportionality assumption is critical to the proportional mean residual life model. In two-arm randomised clinical trials, for example, when the treatment effect is not as persis-

tent as expected, it is possible that there is no constant proportionality in the mean residual life. If situations like this happen, then the proportional mean residual life model needs to be modified to accommodate the time-varying treatment effect, or alternative models should be explored.

An alternative perspective to see that Proposition 1 is true is actually quite intuitive. Since the Weibull family of distribution is the only one to satisfy both the proportional hazards model and the accelerated hazards model, therefore the hazard function of the Hall-Wellner linear mean residual life must satisfy that

$$\frac{1 + D_1}{D_1 t + D_2} \equiv D_3 t^{D_4}.$$

That is, $1 + D_1 \equiv D_1 D_3 t^{D_4+1} + D_2 D_3 t^{D_4}$. Because $D_2 > 0$, $D_1 = D_4 = 0$. As a result, the baseline hazard function must be constant and hence exponential.

To avoid complicate approximation in computing the sandwich estimate of $\hat{\beta}$'s variance-covariance matrix, an alternative resampling approach due to Parzen et al. (1994), which was used also in Lin et al. (1998), can be adapted to estimate the variance-covariance matrix. The basic idea is to perturb the estimating function by a set of simulated independent standard normal deviates, $(e_1, e_2, \ldots, e_n)$:

$$U_e(\beta) = n^{-1} \sum_{i=1}^{n} \frac{\Delta_i}{\hat{F}_c(X_i)} \left\{ Z_i - \bar{Z}_i(\beta, \hat{F}_c) \right\} e_i$$

and solve the solution $\hat{\beta}_e$ of the equation $U_e(\hat{\beta}_e) = e$, where e is also a normal deviate with zero-mean and variance-covariance matrix of $\hat{\Sigma}$. Then according to Parzen, et al. (1994) and Lin, et al. (1998), $n^{1/2}(\hat{\beta} - \hat{\beta}_e)$ has the same limiting distribution as $n^{1/2}(\hat{\beta} - \beta_0)$. Therefore, the variance-covariance matrix of $\hat{\beta}$ can be approximated by the empirical variance-covariance matrix of $\hat{\beta}_e$, given enough large number of sets of $(e_1, e_2, \ldots, e_n)$ and e

are generated.

The estimating function proposed in this article provides one possible way to estimate the parameter in presence of censoring, while treating the baseline mean residual life $m_0(t)$ as nuisance. It is a type of the inverse probability of censoring weighted estimating function. It can be extended to the models in which the censoring satisfies the so-called coarsening-at-random, and furthermore, for instance, the censoring time follows a proportional hazards model:

$$\lambda_c(t|Z) = \lambda_{c,0}(t)\exp(\alpha^{\mathrm{T}} Z).$$

Then the general theory in Robins and Rotnitzky (1992) and van der Laan, Gill and Robins (2000) can be applied to compute its corresponding locally efficient estimating function. Some similar pros and cons of using this "optimal" estimating function can be found in Bang & Tsiatis (2000), although under a different context.

Appendix 1

Proof of Proposition 1

Suppose that there exist β_1 and β_2 such that

$$m(t|Z) = m_0(t)\exp(\beta_1^{\mathrm{T}} Z) = m_0\{t\exp(-\beta_2^{\mathrm{T}} Z)\}\exp(\beta_2^{\mathrm{T}} Z).$$

Since $m_0(t) = \sum_{k=-\infty}^{+\infty} d_k t^k$, therefore

$$\sum_{k=-\infty}^{+\infty} d_k t^k \exp(\beta_1^{\mathrm{T}} Z) = \sum_{k=-\infty}^{+\infty} d_k t^k \exp\{(1-k)\beta_2^{\mathrm{T}} Z\}.$$

Hence, for any pair of k_1 and k_2 such that $d_{k_1} d_{k_2} \neq 0$,

$$\exp(\beta_1^{\mathrm{T}} Z) = \exp\{(1-k_1)\beta_2^{\mathrm{T}} Z\} = \exp\{(1-k_2)\beta_2^{\mathrm{T}} Z\}.$$

As a result, $k_1 = k_2$. That is, $m_0(t)$ is only in the form of $d_k t^k$ for some k. According to the condition (a) in the Hall-Wellner Characterization Theorem (Hall & Wellner, 1981, p. 172), $m_0(t)$ is an appropriate mean residual life when $k = 0$, that is, when the corresponding survival time is exponential. The sufficient condition is straightforward.

Appendix 2

Asymptotic normality of $n^{1/2}U(\beta_0)$

Consider the same martingale representation of $\hat{F}_c$ in §2. Then using the functional Delta-method, we would have

$$\begin{aligned} U(\beta_0) &= n^{-1}\sum_{i=1}^{n}\left\{P_i(\beta_0, F_c) + Q_i(\beta_0, F_c)\frac{F_c(X_i) - \hat{F}_c(X_i)}{F_c(X_i)}\right\} + o_p(n^{-1/2}) \\ &= n^{-1}\sum_{i=1}^{n}\left\{P_i(\beta_0, F_c) + \int_0^{\infty}\frac{\hat{G}(t-)}{G(t)}\frac{A_n(t,\beta_0,F_c)}{B_n(t)}dM_i(t)\right\} + o_p(n^{-1/2}), \quad \text{(A1)} \end{aligned}$$

where

$$P_i(\beta, F_c) = \frac{\Delta_i \left\{ Z_i - \bar{Z}_i(\beta, F_c) \right\}}{F_c(X_i)},$$

and

$$A_n(t, \beta, F_c) = n^{-1} \sum_{i=1}^{n} Q_i(\beta, F_c) I(X_i \geq t) \to A(t, \beta, F_c).$$

Here $Q_i(\beta, F_c)$ is computed as follows:

$$\begin{aligned}
Q_i(\beta, F_c) &= -F_c \frac{\partial}{\partial F_c} \left[\frac{\Delta_i}{F_c} \left\{ Z_i - \bar{Z}_i(\beta, F_c) \right\} \right] \\
&= -F_c \left[\frac{\partial}{\partial F_c} \left(\frac{\Delta_i}{F_c} \right) \left\{ Z_i - \bar{Z}_i(\beta, F_c) \right\} + \frac{\Delta_i}{F_c} \frac{\partial}{\partial F_c} \left\{ Z_i - \bar{Z}_i(\beta, F_c) \right\} \right] \\
&= -F_c \left[-\frac{\Delta_i \left\{ Z_i - \bar{Z}_i(\beta, F_c) \right\}}{F_c^2} - \frac{\Delta_i}{F_c} \frac{\partial \bar{Z}_i(\beta, F_c)}{\partial F_c} \right] \\
&= P_i(\beta, F_c) + R_i(\beta, F_c),
\end{aligned}$$

where

$$\begin{aligned}
&R_i(\beta, F_c) \\
&= \Delta_i \int_0^\infty \frac{\partial}{\partial F_c} \left[\frac{\sum_i Z_i \exp(-2\beta^{\mathrm{T}} Z_i)(X_i - t)^+ \Delta_i \{F_c(X_i)\}^{-1}}{\sum_i \exp(-2\beta^{\mathrm{T}} Z_i)(X_i - t)^+ \Delta_i \{F_c(X_i)\}^{-1}} \right. \\
&\left. \times \frac{\exp(-\beta^{\mathrm{T}} Z_i) I(X_i > t)}{n^{-1} \sum_i \exp(-\beta^{\mathrm{T}} Z_i) X_i \Delta_i \{F_c(X_i)\}^{-1}} \right] dt \\
&= \int_0^\infty \frac{K_{i1}(\beta, F_c) \exp(-\beta^{\mathrm{T}} Z_i) I(X_i > t)}{n^{-1} \sum_i \exp(-\beta^{\mathrm{T}} Z_i) X_i \Delta_i \{F_c(X_i)\}^{-1}} dt \\
&+ \int_0^\infty \frac{K_{i2}(\beta, F_c) \sum_i Z_i \exp(-2\beta^{\mathrm{T}} Z_i)(X_i - t)^+ \Delta_i \{F_c(X_i)\}^{-1}}{\sum_{i=1}^n \exp(-2\beta^{\mathrm{T}} Z_i)(X_i - t)^+ \Delta_i \{F_c(X_i)\}^{-1}} dt,
\end{aligned}$$

$$
\begin{aligned}
&K_{1i}(\beta, F_c) \\
&= \frac{\partial}{\partial F_c}\left[\frac{\sum_i Z_i \exp(-2\beta^{\mathrm{T}} Z_i)(X_i - t)^+ \Delta_i \{F_c(X_i)\}^{-1}}{\sum_i \exp(-2\beta^{\mathrm{T}} Z_i)(X_i - t)^+ \Delta_i \{F_c(X_i)\}^{-1}}\right] \\
&= \left[\sum_{i=1}^n \exp(-2\beta^{\mathrm{T}} Z_i)(X_i - t)^+ \Delta_i \{F_c(X_i)\}^{-1}\right]^{-2} \\
&\times \left\{\left[\sum_{i=1}^n Z_i \exp(-2\beta^{\mathrm{T}} Z_i)(X_i - t)^+ \Delta_i \{F_c(X_i)\}^{-1}\right]\left[\sum_{i=1}^n \exp(-2\beta^{\mathrm{T}} Z_i)(X_i - t)^+ \Delta_i \{F_c(X_i)\}^{-2}\right]\right. \\
&\left. - \left[\sum_{i=1}^n Z_i \exp(-2\beta^{\mathrm{T}} Z_i)(X_i - t)^+ \Delta_i \{F_c(X_i)\}^{-2}\right]\left[\sum_{i=1}^n \exp(-2\beta^{\mathrm{T}} Z_i)(X_i - t)^+ \Delta_i \{F_c(X_i)\}^{-1}\right]\right\},
\end{aligned}
$$

and

$$
\begin{aligned}
&K_{2i}(\beta, F_c) \\
&= \frac{\partial}{\partial F_c}\left[\frac{\exp(-\beta^{\mathrm{T}} Z_i) I(X_i > t)}{n^{-1}\sum_i \exp(- = \beta^{\mathrm{T}} Z_i) X_i \Delta_i \{F_c(X_i)\}^{-1}}\right] \\
&= \left\{\exp(-\beta^{\mathrm{T}} Z_i) I(X_i > t)\right\} \\
&\times \left[n^{-1}\sum_{i=1}^n \exp(-\beta^{\mathrm{T}} Z_i) X_i \Delta_i \{F_c(X_i)\}^{-1}\right]^{-2}\left[n^{-1}\sum_{i=1}^n \exp(-\beta^{\mathrm{T}} Z_i) X_i \Delta_i \{F_c(X_i)\}^{-2}\right],
\end{aligned}
$$

Then it follows a Multivariate Central Limit Theorem that $n^{1/2}U(\beta_0)$ is asymptotically normal with mean zero and variance and covariance matrix

$$
\Sigma\left(\beta_0, F_c\right) = E\left[\left\{P(\beta_0, F_c) + \int_0^\infty \frac{A(t, \beta_0, F_c)}{B(t)} dM_1(t)\right\}^{\otimes 2}\right]
$$

Here $v^{\otimes 2} = vv^{\mathrm{T}}$. Furthermore, if Z_i are assumed to be bounded, since $Z_i - \bar{Z}(\beta, F_c)$ is $\mathcal{F}(0)$-measurable, the first two terms in (A1) is uncorrelated. It is therefore follows standard variance calculation for martingales that

$$
\Sigma = E\left\{P(\beta_0, F_c)\right\}^{\otimes 2} + \int_0^\infty \frac{A(t, \beta_0, F_c) A^{\mathrm{T}}(t, \beta_0, F_c)}{B(t)} d\Lambda_c(t)
$$

References

Bang, H. & Tsiatis, A. A. (2000) Estimating medical cost with censored data. *Biometrika* **87**, 329-43.

Chiang, C. L. (1960) A stochastic study of the life table and its applications: I. Probability distributions of the biometric functions. *Biometrics* **16**, 618-35.

Cox, D. R. (1962) *Renewal Theory.* New York: Wiley.

Cox, D. R. (1972) Regression models and life-tables (with discussion). *J. R. Statist. Soc.* B **34**, 187-220.

Csörgő, M. & Zitikis, R. (1996) Mean residual life processes. *Ann. Statist.* **4**, 1717-39.

Fleming, T. R., & Harrington, D. P. (1991) *Counting Processes and Survival Analysis.* New York: John Wiley.

Gill, R. D. & Schumacher, M. (1987). A simple test of the proportional hazards assumption. *Biometrika* **74**, 289-300.

Guess, F. & Proschan, F. (1988) Mean residual life: theory and application. In *Handbook of Statistics: Quality Control and Reliability* (P. R. Krishnaiah and C. R. Rao, eds.) **7**. Amsterdam: North-Holland.

Gupta, R. C., & Kirmani, S. (1998). On the proportional mean residual life model and its implications. *Statist.* **32**, 175-87.

Kalbfleisch, J. D. & Prentice, R. L. (1980) *The Statistical Analysis of Failure Time Data.* New York: Wiley.

LAD, T., RUBINSTEIN L. & SADEGHI A. (1988) The benefit of adjuvant treatment for resected locally advanced non-small-cell lung cancer. *J. Clin. Oncol.* **6**, 9-17.

LIN, D. Y. (1991). Goodness of fit for the Cox regression model based on a class of parameter estimators. *J. Am. Statist. Assoc.* **86**, 725-8.

LIN, D. Y., WEI, L. J. & YING, Z. (1998) Accelerated failure time models for counting processes. *Biometrika* **85**, 605-18.

MAGULURI, G. & ZHANG, C.-H. (1994) Estimation in the mean residual life regression model. *J. R. Statist. Soc.* B **56**, 477-89.

MEILIJSON, I. (1972). Limiting properties of the mean residual life function. *Ann. Math. Statist.* **43**, 354-57.

OAKES, D. & DASU, T. (1990) A note on residual life. *Biometrika* **77**, 409-10.

PARZEN, M. I., WEI, L. J. & YING, Z. (1994). A resampling method based on pivotal estimating functions. *Biometrika* **81**, 341-50.

PIANTADOSI, S. (1997). *Clinical Trials: A Methodologic Perspective.* New York: John Wiley.

VAN DER LAAN, M. J., GILL, R. D. & ROBINS, J. M. (2000) Locally efficient estimation in censored data models: theory and examples. *UC Berkeley Biostatistics Technical Report*, 85.

ZAHEDI, H. (1991) Proportional mean remaining life model. *J. Statist. Plan. and Inf.* **29**, 221-28.

Table 1: Summary of Simulation Studies

			Rescaled Beta			Exponential			Pareto		
n	Cens.		$\beta_0 = -1$	$\beta_0 = 0$	$\beta_0 = 1$	$\beta_0 = -1$	$\beta_0 = 0$	$\beta_0 = 1$	$\beta_0 = -1$	$\beta_0 = 0$	$\beta_0 = 1$
100	0%	Bias	0.009	-0.001	0.003	0.007	0.002	-0.001			
100	0%	Cov.	0.952	0.948	0.954	0.952					

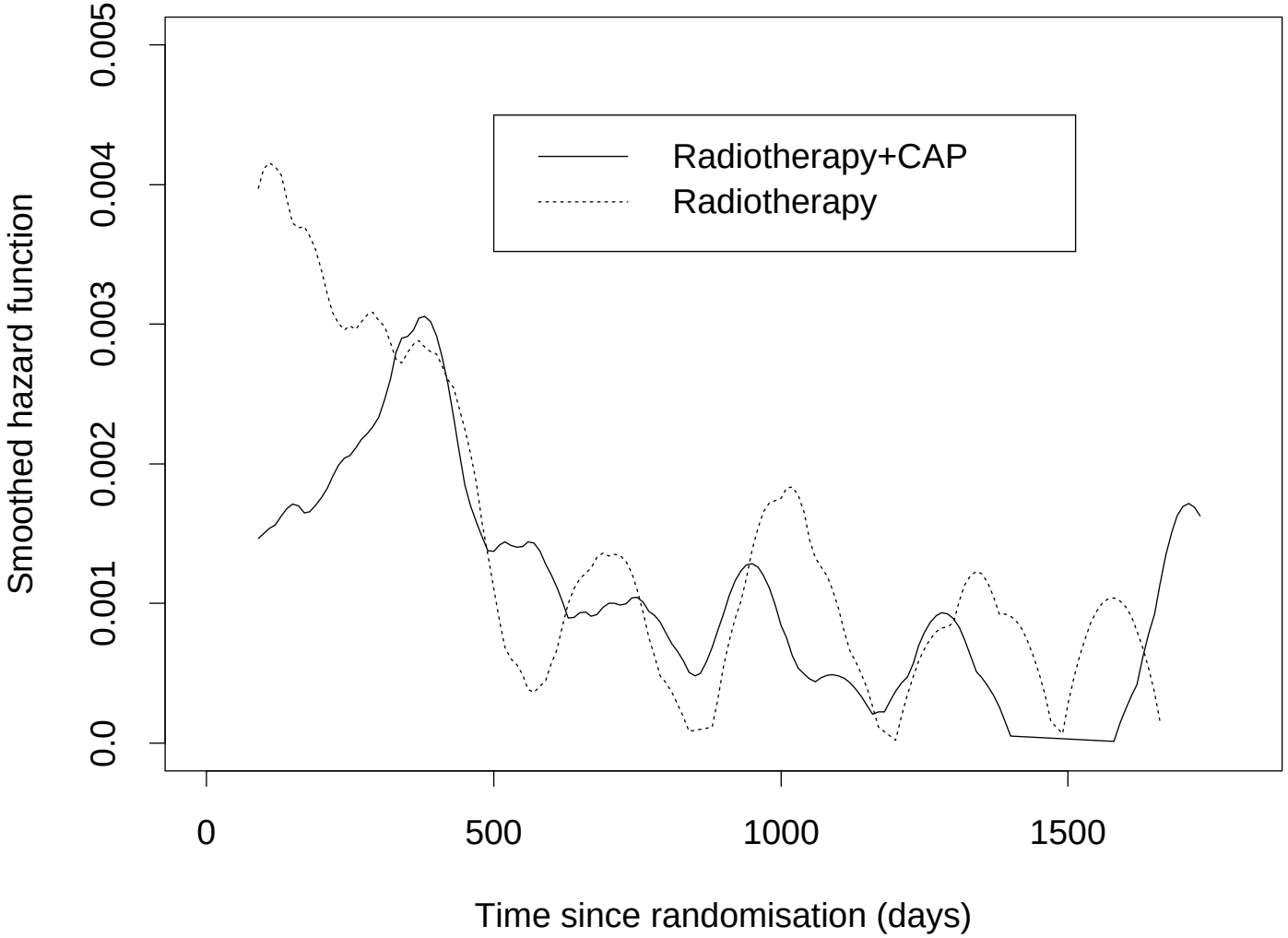

Figure 1: Smoothed hazard functions in comparative treatment efficacy lung cancer trial
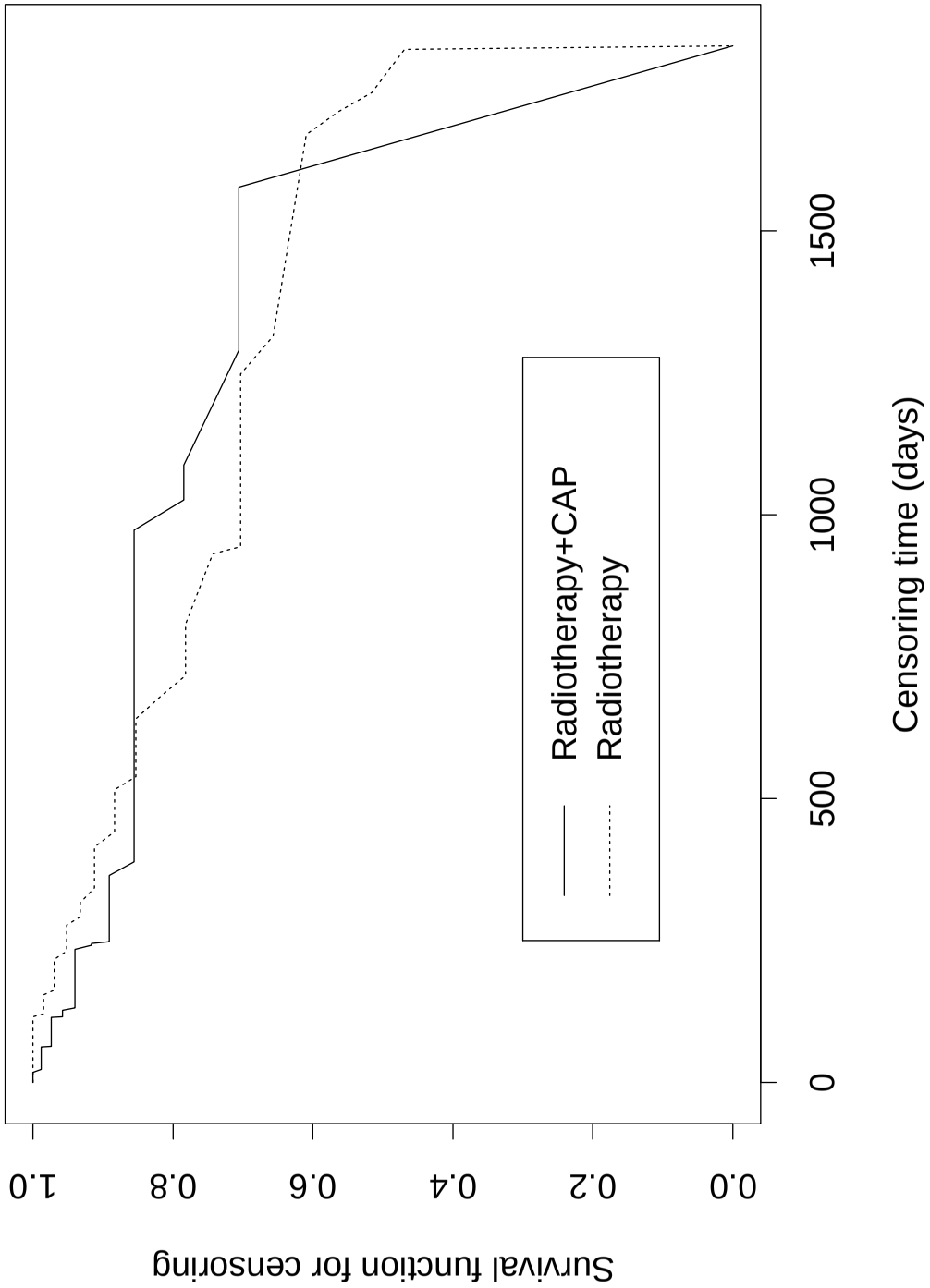
Figure 2: Survival functions for censoring times in COMPARATIVE TREATMENT EFFICACY lung cancer trial