Confidence Intervals for the Difference of Median Survival Times Using the Stratified Cox Proportional Hazards Model

JINHEUM KIM

Department of Applied Statistics
University of Suwon
Suwon P.O. Box 77
Kyeonggi-Do, 445-743
South Korea

Summary

The stratified Cox proportional hazards model is introduced to incorporate covariates and involve non-proportional treatment effect of two groups into the analysis and then the confidence interval estimators for the difference in median survival times of two treatments in stratified Cox model are proposed. The one is based on baseline survival functions of two groups, and the other on average survival functions of two groups. I illustrate the proposed methods with an example from a study conducted by the Radiation Therapy Oncology Group in cancer of the mouth and throat. Simulations are carried out to investigate the small-sample properties of proposed methods in terms of coverage rates.

Key words: Censored data; Confidence interval; Median survival time; Stratified Cox model.

1. Introduction

In survival analysis, the median survival time is frequently used as a measure for comparing effects on survival of two treatments. In the presence of covariates information, the Cox proportional hazards model may be easily considered to investigate how covariates affect the survival and thereby the median survival time. In the Cox model, it is assumed that the ratio of effects of two treatments on hazard rate is constant over time. As mentioned in KARRISON (1987, 1997) and ZUCKER (1998), when the treatment effects of two groups conflict around some time \( t_0 \), i.e., when one treatment is better than the other up to time \( t_0 \) but subsequently the relation is reversed, the Cox model no longer reflects this situation. To incorporate nonproportional treatment effect of two groups into a model, KARRISON (1987, 1997) introduced a piecewise exponential model with proportional hazards covariate effects, and ZUCKER (1998) considered a stratified Cox proportional hazards model less subjective than that of KARRISON (1987, 1997) in a
viewpoint of no need to break up time interval into arbitrary subintervals.

In this paper, as in Zucker (1998), I incorporate covariates into the analysis through the stratified Cox proportional hazards model with group being considered as a stratification variable. The model specifies that the hazard rate for the survival time of a subject with covariate vector $x' = (x_1, \ldots, x_p)$ in group $i$, say $i = 1, 2$, is

$$
\lambda_i(t \mid x) = \lambda_{0i}(t) \exp \left( \beta'_i x \right),
$$

where $\lambda_{0i}(\cdot)$ is an unspecified baseline hazard function and $\beta_i$ is a $p \times 1$ vector of unknown regression coefficients.

Section 2 introduces notation to be mentioned in the followings and reviews basic results related with the stratified Cox model. In Sections 3 and 4, approximate confidence interval estimates for the difference in median survival times based on baseline survival functions and average survival functions of two groups are derived respectively. Section 5 contains an example from a study conducted by the Radiation Therapy Oncology Group in cancer of the mouth and throat. In Section 6, simulations are carried out to investigate the small-sample properties of proposed methods in terms of coverage rates.

2. Notation and Review of Basic Results

Let $T_{i1}, \ldots, T_{in_i}$ be independent survival times in the $i$th group ($i = 1, 2$) and $X_{i1}, \ldots, X_{in_i}$ be the corresponding covariate vectors. Suppose that given $X_{ij} = x_{ij}$, the $T_{ij}$ follows model (1). Let $C_{ij}$ be the potential censoring times for $T_{ij}$. The $T_{ij}$’s are subject to right censoring so that we observe $(Y_{ij}, \delta_{ij}, X_{ij})$ ($i = 1, 2; j = 1, \ldots, n_i$), where $Y_{ij} = T_{ij} \wedge C_{ij}$, $\delta_{ij} = I(T_{ij} \leq C_{ij})$. Assume that $T$ and $C$ are conditionally independent given $X$.

For $i = 1, 2; j = 1, \ldots, n_i$, define the processes $N_{ij}(t) = I(Y_{ij} \leq t, \delta_{ij} = 1)$ and $J_{ij}(t) = I(Y_{ij} \geq t)$. Following Andersen et al. (1993), for $i = 1, 2$, let

$$
S_i^{(0)}(\beta, t) = \sum_{j=1}^{n_i} J_{ij}(t) \exp \left( \beta' X_{ij} \right), \quad S_i^{(1)}(\beta, t) = \sum_{j=1}^{n_i} X_{ij} J_{ij}(t) \exp \left( \beta' X_{ij} \right),
$$

$$
S_i^{(2)}(\beta, t) = \sum_{j=1}^{n_i} X_{ij}^2 J_{ij}(t) \exp \left( \beta' X_{ij} \right),
$$

$$
E_i(\beta, t) = S_i^{(0)}(\beta, t)^{-1} S_i^{(1)}(\beta, t),
$$

$$
V_i(\beta, t) = S_i^{(0)}(\beta, t)^{-1} S_i^{(2)}(\beta, t) - E_i(\beta, t)^{\otimes 2},
$$

where for a $p \times 1$ vector $a$, $a^{\otimes 2}$ is the $p \times p$ matrix $aa'$. Assume Conditions VII.2.1 and VII.2.2 of Andersen et al. (1993, p. 497) and

$$
\int_0^t \lambda_{0i}(s) \, ds < \infty,
$$

(2)
for \( i = 1, 2 \) and \( t \in [0, \tau] \), the time interval over which the subjects are observed. In particular, Condition (a) of Condition VII.2.1 states that for \( i = 1, 2 \) and \( k = 0, 1, 2, \) \( n^{-1}S_i^{(k)}(\mathbf{b}, t) \) converges in probability to \( s_i^{(k)}(\mathbf{b}, t) \) uniformly in \( t \in [0, \tau] \) and in a neighborhood of \( \mathbf{b}_0 \). For \( i = 1, 2 \), let
\[
e_i(\mathbf{b}, t) = s_i^{(0)}(\mathbf{b}, t)^{-1} s_i^{(1)}(\mathbf{b}, t),
\]
\[
v_i(\mathbf{b}, t) = s_i^{(0)}(\mathbf{b}, t)^{-1} s_i^{(2)}(\mathbf{b}, t) - e_i(\mathbf{b}, t)^{\otimes 2}.
\]
Define
\[
\Sigma = \sum_{i=1}^{2} \int_0^{\tau} v_i(\mathbf{b}_0, s) s_i^{(0)}(\mathbf{b}_0, s) \lambda_{o_i}(s) \, ds,
\]
and assume that \( \Sigma \) is positive definite. From Andersen et al. (1993), for the maximum partial likelihood estimator \( \hat{\mathbf{b}} \) of \( \mathbf{b}_0 \), \( n^{1/2}(\hat{\mathbf{b}} - \mathbf{b}_0) \) is asymptotically normal with mean 0 and covariance \( \Sigma^{-1} \), where \( n = n_1 + n_2 \). Define the cumulative baseline hazard function of group \( i (i = 1, 2) \) and its corresponding estimate by
\[
\Lambda_{o_i}(t) = \int_0^t \lambda_{o_i}(s) \, ds, \quad \hat{\Lambda}_{o_i}(t) = \int_0^t s_i^{(0)}(\hat{\mathbf{b}}, s)^{-1} dN_i(s),
\]
where \( N_i(t) = \sum_j N_{ij}(t) \) \( (i = 1, 2) \).

3. Approach Based on Baseline Survival

Let \( S_{o_i}(t) \) denote the baseline survival function for group \( i (i = 1, 2) \). Then, \( S_{o_i}(t) = \exp\{-\Lambda_{o_i}(t)\} \) and \( S_{o_i}(t) \) is estimated by \( \hat{S}_{o_i}(t) = \exp\{-\hat{\Lambda}_{o_i}(t)\} \). Using Corollary VII.2.4 of Andersen et al. (1993), note that \( [n^{1/2}(\hat{S}_{o_i}(\cdot) - S_{o_i}(\cdot)), i = 1, 2] \) converges weakly to a 2-variate Gaussian process with mean zero and covariance function
\[
\sigma(s, t) = S_{o_i}(s) S_{o_i'}(t) \{\delta_{ii'} a_i(s \wedge t) + b_i(s) b_i'(t)\Sigma^{-1} b_i'(t)\}, \quad i, i' = 1, 2,
\]
where \( \delta_{ii'} = 1 \) if \( i = i' \), and 0, otherwise and
\[
a_i(t) = \int_0^t \{s_i^{(0)}(\mathbf{b}_0, s)\}^{-1} \lambda_{o_i}(s) \, ds, \quad b_i(t) = \int_0^t e_i(\mathbf{b}_0, s) \lambda_{o_i}(s) \, ds.
\]

Let \( 0 < p < 1 \), and define the \( p \)th quantile of survival time \( T \) for group \( i \) based on \( S_{o_i} \) as
\[
\tilde{\xi}_i^{(p)} = \inf \{t : S_{o_i}(t) \leq 1 - p\}, \quad i = 1, 2.
\]
Then, its natural estimate, say \( \hat{\xi}_i^{(p)} \), is obtained by substituting \( \hat{S}_{o_i} \) for \( S_{o_i} \). Also, define the difference between \( p \)th quantiles of each group as
\[
\Delta = \hat{\xi}_1^{(p)} - \hat{\xi}_2^{(p)}.
\]
and its corresponding estimate as \( \hat{\Delta} = \hat{\xi}_1^{(p)} - \hat{\xi}_2^{(p)} \). Set
\[
\Theta(t_1, t_2) = \{\lambda_{01}(t_1)\}^{-1} \mathbf{b}_1(t_1) - \{\lambda_{02}(t_2)\}^{-1} \mathbf{b}_2(t_2).
\]

**Theorem 1:** Assume Conditions VII.2.1 and VII.2.2 of Andersen et al. (1993) and that (2) holds for \( t = \tau \). In addition, assume that \( 0 < p_1 < p_2 < 1 \) are points such that \( \lambda_{0i}(t) \) \( (i = 1, 2) \) is continuous and bounded away from zero for \( t \in [\xi_i^{(p_1)}, \xi_i^{(p_2)}] \). Then, \( n^{1/2}(\Delta - \Delta) \) converges weakly on \( D[p_1, p_2] \) to a zero-mean Gaussian process with variance function
\[
\sigma^2(\xi_i^{(p)}, \xi_{i'}^{(p)}) = \sum_{i=1}^{2} \{\lambda_{0i}^2(\xi_i^{(p)})\}^{-1} a_i(\xi_i^{(p)}) + \Theta(\xi_i^{(p)}, \xi_{i'}^{(p)})' \Sigma^{-1} \Theta(\xi_i^{(p)}, \xi_{i'}^{(p)}),
\]
where \( D[a, b] \) denotes the space of functions on \( [a, b] \) which are right-continuous with left-hand limit.

**Proof:** From Theorem 1 of Doss and Gill (1992), \( \{n^{1/2}(\hat{\xi}_i^{(p)} - \xi_i^{(p)}), i = 1, 2\} \) is asymptotically equivalent to
\[
[n^{1/2}\{1 - p\} \lambda_{0i}(\xi_i^{(p)})]^{-1} \{\hat{S}_{0i}(\xi_i^{(p)}) - S_{0i}(\xi_i^{(p)})\}, i = 1, 2,
\]
and thereby, from weak convergence of \( [n^{1/2}\{\hat{S}_{0i}(\cdot) - S_{0i}(\cdot)\}, i = 1, 2] \), converges weakly to a 2-variate Gaussian process with mean zero and covariance function
\[
\{(1 - p)^2 \lambda_{0i}(\xi_i^{(p)}) \lambda_{0i'}(\xi_{i'}^{(p)})\}^{-1} \sigma(\xi_i^{(p)}, \xi_{i'}^{(p)}), i, i' = 1, 2.
\]
Thus, we immediately hold the theorem. \( \square \)

To estimate \( \sigma \), it is necessary to estimate the functions \( a_i, b_i, \lambda_{0i} \), and \( \Sigma \). At first, for \( a_i, b_i, \) and \( \Sigma \), one may introduce the natural estimates
\[
\hat{a}_i(t) = n \int_0^t \{ S_i^{(0)}(\hat{\mathbf{b}}, s) \}^{-1} d\hat{\Lambda}_{0i}(s), \quad \hat{b}_i(t) = \int_0^t E_i(\hat{\mathbf{b}}, s) d\hat{\Lambda}_{0i}(s),
\]
\[
\hat{\Sigma} = n^{-1} \sum_{i=1}^{2} \int_0^\tau V_i(\hat{\mathbf{b}}, s) S_i^{(0)}(\hat{\mathbf{b}}, s) d\hat{\Lambda}_{0i}(s),
\]
respectively. Secondly, estimating \( \lambda_{0i} \) is based on a smoothing procedure using a kernel function proposed by Wells (1994). To describe them, for \( i = 1, 2 \), let \( K_i \) be a function of bounded variation with support on \( [-1, 1] \) and whose integral is 1, and \( h_i \) depending on \( n \) be the bandwidth such that as \( n \to \infty \), \( h_i \to 0 \) and \( nh_i \to \infty \). Define the kernel estimate of \( \lambda_{0i}(\cdot) \) by
\[
\hat{\lambda}_{0i}(t) = h_i^{-1} \int_0^\infty K_i \left( \frac{t - s}{h_i} \right) d\hat{\Lambda}_{0i}(s).
\]
It has been shown by Wells (1994) that \( \hat{\theta}_0i \) is uniformly consistent. Having specified a choice of \( K_i \) and \( h_i \), one may define an estimate \( \hat{\nu} \) of \( \nu \) by substituting in (3) \( \hat{a}_i, \hat{b}_i, \Sigma, \hat{\xi}_i^{(p)} \), and \( \hat{\lambda}_0i \) for \( a_i, b_i, \Sigma, \xi_i^{(p)}, \) and \( \lambda_0i \). Since each estimate in \( \hat{\nu} \) is consistent, it is easy to see that \( \hat{\nu} \to \nu \) in probability as \( n \to \infty \). Thus, an approximate 100(1 − \( \alpha \))\% confidence interval for \( \Delta \) is given by

\[
\hat{\Delta} \pm z_{\frac{1}{2}\alpha} n^{-\frac{1}{2}} \hat{\nu}^2, \tag{4}
\]

where \( z_{\frac{1}{2}\alpha} \) is the \( (1 - \frac{1}{2}\alpha)\)th quantile of the standard normal distribution.

4. Approach Based on Average Survival

Let \( S_i(t \mid x) \) denote the survival function for a subject with covariate vector \( X = x \) in group \( i \) (\( i = 1, 2 \)). Under model (1), \( S_i(t \mid x) = \exp \{ -\exp (\beta'_i x) \Lambda_0i(t) \} \), and its natural estimate is given by \( \hat{S}_i(t \mid x) = \exp \{ -\exp (\hat{\beta}' x) \Lambda_0i(t) \} \). Define the average survival function for group \( i \), as in Zucker (1998), as

\[
S_i(t) = n^{-1} \sum_{g=1}^{n_g} \sum_{j=1}^{n_g} S_i(t \mid x_{gi}), \quad i = 1, 2,
\]

and the \( p \)th quantile of group \( i \) based on the average survival function \( S_i \) as

\[
\hat{\xi}^{(p)} = \inf \{ t : S_i(t) \leq 1 - p \}, \quad i = 1, 2.
\]

Then, their respective natural estimates are obtained by substituting \( \hat{S}_i(\cdot \mid x) \) for \( S_i(\cdot \mid x) \), and thereby \( \hat{S}_i \) for \( S_i \). In addition, define the difference between \( p \)th quantiles of each group as

\[
\Delta_i = \hat{\xi}^{(p)} - \hat{\xi}^{(p)}_2,
\]

and its corresponding estimate as \( \hat{\Delta}_i = \hat{\xi}^{(p)} - \hat{\xi}^{(p)}_2 \). For simplicity of notation, let, for \( i = 1, 2, \)

\[
\Gamma_i(t) = n^{-1} \sum_{g=1}^{n_g} \sum_{j=1}^{n_g} S_i(t \mid x_{gi}) \exp (\hat{\beta}'_i x_{gi}),
\]

\[
\Omega_i(t) = n^{-1} \sum_{g=1}^{n_g} \sum_{j=1}^{n_g} S_i(t \mid x_{gi}) \exp (\hat{\beta}'_i x_{gi}) x_{gi},
\]

\[
\Phi_i(t) = -\Lambda_0i(t) \Omega_i(t) + \Gamma_i(t) b_i(t),
\]

and let

\[
\Psi(t_1, t_2) = \{\lambda_{01}(t_1) \Gamma_1(t_1)\}^{-1} \lambda_{01}(t_1) \Omega_1(t_1) - \{\lambda_{02}(t_2) \Gamma_2(t_2)\}^{-1}
\]

\[
\times \lambda_{02}(t_2) \Omega_2(t_2).
\]
Using the arguments in Sections VII.2.2 and VII.2.3 of Andersen et al. (1993), note that \( n^2 \{ \hat{S}_i(t) - S_i(t) \}, i = 1, 2 \) converges weakly to a 2-variate Gaussian process with mean zero and covariance function

\[
\tau(s, t) = \delta_{ii} \Gamma_i(s) \Gamma_i(t) a_i(s \land t) + \Phi_i(s)' \Sigma^{-1} \Phi_i(t), \quad i, i' = 1, 2.
\]

Therefore, using the arguments of Theorem 1, we have the following theorem similar to Theorem 1.

Theorem 2: Under the same assumptions in Theorem 1 except replacing \( \xi_i^{(p)} \) by \( \xi_i^{(p)} \), \( n^2 (\Delta - \Delta) \) converges weakly on \( D[p_1, p_2] \) to a zero-mean Gaussian process with variance function

\[
\nu(\xi_{i1}^{(p)}, \xi_{i2}^{(p)}) = \sum_{i=1}^{2} \{ \lambda_{0i}(\xi_{i}^{(p)}) \}^{-1} a_i(\xi_{i}^{(p)}) + \Theta(\xi_{i1}^{(p)}, \xi_{i2}^{(p)}) - \Psi(\xi_{i1}^{(p)}, \xi_{i2}^{(p)}) \}' \times \Sigma^{-1} \{ \Theta(\xi_{i1}^{(p)}, \xi_{i2}^{(p)}) - \Psi(\xi_{i1}^{(p)}, \xi_{i2}^{(p)}) \}.
\]

Furthermore, it follows as in Section 3 that a consistent estimate \( \hat{\nu} \) of \( \nu \) can be obtained by replacing in (7) \( \hat{a}_i, \hat{b}_i, \hat{\Sigma}, \hat{\xi}_i^{(p)}, \hat{\lambda}_{0i}, \hat{S}_i(\cdot | x), \hat{\beta}, \) and \( \hat{\Lambda}_{0i} \) for \( a_i, b_i, \Sigma, \xi_i^{(p)}, \lambda_{0i}, S_i(\cdot | x), \beta_0, \) and \( \Lambda_{0i} \). Thus, an approximate 100(1 - \( \alpha \))% confidence interval for \( \Delta \) is given by

\[
\hat{\Delta} \pm z_{\alpha/2} n^{-1/2} \hat{\nu}.
\]

5. An Example

I illustrate the proposed methods with a dataset in Appendix 1 of Kalbfleisch and Prentice (1980). This dataset is a subset of the data from a randomized clinical trial conducted by the Radiation Therapy Oncology Group to compare radiation therapy (\( R \)) alone versus chemotherapy followed by radiation therapy (\( C + R \)) in the treatment of carcinoma of the mouth and throat. The dataset includes survival time in days and status for 195 patients that consists of 100 patients who received only \( R \) and 95 patients who received \( C + R \), along with 6 covariates such as sex, tumor grade, tumor site, general condition, \( T \) stage and \( N \) stage. All the covariates were centered at their overall mean values in combined groups. For details, see Karrison (1987) and Zucker(1998). To estimate baseline hazards, Epanechnikov kernel function, defined by \( K(t) = 0.75(1 - |t|^2) I(|t| \leq 1) \), was used and bandwidth was chosen using the maximum likelihood cross-validation (ML-CV) method (see Härdle (1991), pp. 93–95). Figure 1 displays kernel estimates of baseline hazard rates of \( R \) only and \( C + R \) groups corresponding to respective optimal bandwidths 230 and 335 in days. It is observed from Figure 1 that the treatment effects of two groups are nonproportional rather than proportional because two curves are crossing around 630 days. The spike at the end of
the $R$ only hazard rate curve is due to the low precision in the tail. However, this does not affect an effect on the estimate of median. For $p = 0.5$, based on baseline survivals, $\hat{\Delta} = 73(\pm 82)$ days with quantiles 480 days for $R$ only group and 407 days for $C + R$ one, and approximate 95% confidence interval for $\Delta$ is $(-88, 234)$. Based on average survivals, $\hat{\Delta} = 76(\pm 80)$ days with quantiles 480 days for $R$ only group and 404 days for $C + R$ one, and approximate 95% confidence interval for $\Delta$ is $(-81, 233)$. As shown, the results based on baseline and average survival curves are so similar. This is why the covariates were centered at their overall mean values. These results implies that $R$ only group is a little better than $C + R$ one in terms of median survival times. Also, it may be stated from Figures 2 and 3 that at early time group effect does not exist but difference between groups gradually exposes passing through about 270 days and eventually disappears.
6. Simulation Studies

Simulations were performed to investigate the small-sample properties of interval estimators (4) and (6). The assumed survival distribution is a Weibull of the form

\[ S_0(t) = \exp \left\{-\left(\alpha_i t\right)^{\gamma_i}\right\}, \quad i = 1, 2, \]

where \( \alpha_i > 0 \) and \( \gamma_i > 0 \) are the scale and shape parameters, respectively. As in Karrison (1997), four configurations of Weibull parameter values, i.e., \((\alpha_1, \alpha_2, \gamma_1, \gamma_2)\), were chosen to cover a few characteristic situations as follows: (i) cases without and with a group effect in proportional hazards (ii) cases with early and late group effects in nonproportional hazards. In each configuration, a total of 3000 simulations were performed and a set of single covariate values of size \( n \) generated from uniform \([0,1]\) was used for all the simulations. The sample sizes of each group are equal and the true regression parameter \( \beta_0 \) is 0.4. The censoring distribution is uniform \([6,8]\). Under Weibull survival distribution, the true value of \( \Delta \) is equal to \( \Delta = \alpha_1^{-1}\{-\ln (1 - p)\}^{1/\gamma_1} - \alpha_2^{-1}\{-\ln (1 - p)\}^{1/\gamma_2} \), and that of \( \Delta \) can be uniquely obtained by solving a numerical equation under assumption that an observed sample of covariates is fixed in each configuration. To estimate baseline hazards, as in Section 5, Epanechnikov kernel was used and optimal bandwidth of each group was chosen using the ML-CV method.

Table 1 presents the sample statistics of \( \hat{\Delta} \) and \( \hat{\Delta} \), and the empirical coverage probabilities of confidence intervals, (4) and (6), for the difference in median survival times of two groups. Four configurations, say, I, II, III, and IV, correspond to \((\alpha_1, \alpha_2, \gamma_1, \gamma_2) = (0.18, 0.18, 1.25, 1.25), (0.14, 0.22, 1.25, 1.25), (0.16, 0.18, 1.50, 0.75), \) and \((0.16, 0.26, 1.25, 1.75)\), respectively. The average censoring fractions for configurations, I, II, III, and IV, are about 20%, 22%, 23%, and 14%, respectively. In simulations there were instances of undefined medians, in particular, when the sample size is small. Whenever it happened, the sample was re-
placed by the new one. From the entries of sample statistics of $\hat{\Delta}$ and $\hat{\Delta}$, in Table 1, a mean squared error decreases as the sample size increases and also normal approximations of $\hat{\Delta}$ and $\hat{\Delta}$ work well in small-sample by means of comparison between SD and SE. Table 1 implies that the empirical coverage probabilities substantially achieve the nominal levels regardless of four different configurations. Also, the coverage rates between configurations are more stable in large sample than in small one.

### Acknowledgements

The author is grateful to two anonymous referees and Jianwen Cai for many valuable comments and useful suggestions on this article. This work was supported by Korea Research Foundation Grant (KRF-2001-015-DP0070).
References


JINHEUM KIM
Department of Applied Statistics
University of Suwon
Suwon P.O. Box 77
Kyeonggi-Do, 445-743
South Korea
Email: jinhkim@mail.suwon.ac.kr; jkim@bios.unc.edu

Received, September 2000
Revised, February 2001
Accepted, April 2001