HHS Public Access

Author manuscript

*J Biopharm Stat*. Author manuscript; available in PMC 2019 January 01.

Published in final edited form as:

*J Biopharm Stat*. 2018 ; 28(5): 927–938. doi:10.1080/10543406.2017.1405012.

# Statistical Methods for Conditional Survival Analysis

**Sin-Ho Jung**[a], **Ho Yun Lee**[b], and **Shein-Chung Chow**[a]

[a]Department of Biostatistics and Bioinformatics, Duke University, Durham, NC, USA;

[b]Department of Radiology, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Korea

## SUMMARY

We investigate the survival distribution of the patients who have survived over a certain time period. This is called a conditional survival distribution. In this paper, we show that one-sample estimation, two-sample comparison and regression analysis of conditional survival distributions can be conducted using the regular methods for unconditional survival distributions that are provided by the standard statistical software, such as SAS and SPSS. We conduct extensive simulations to evaluate the finite sample property of these conditional survival analysis methods. We illustrate these methods with real clinical data.

### Keywords

Delta method; Fieller method; Kaplan-Meier estimator; Log-rank test; Martingale central limit theorem; Proportional hazards model

## 1 Introduction

Traditionally, survival analysis in clinical researches has been to investigate the distribution of patients' survival times measured from the diagnosis of a disease or the start of a treatment (i.e., baseline). This type of analysis provides the survival probability of patients expected at the start of treatment that will be useful to predict their prognosis before starting the treatment. However, the survival probability evolves over time and usually decrease with increased survivorship, so that both patients and clinicians are interested in the change in survival probability over the progress of treatment and disease. Residual lifetime of individuals when they have survived over a relevant landmark of time can serve towards this end. For example, when comparing the efficacy of an intensive treatment with a standard treatment, patients receiving the prior may have a higher risk due to treatment-related mortality during treatment period, but may have a much lower risk once they survive over the treatment period with the disease cured.

While researches on residual lifetime theory have been very active in reliability area (e.g. Bryson and Siddiqui, 1969; Hollander and Proschan, 1975; Muth, 1977; Ruiz and Navarro,

Correspondence: Sin-Ho Jung, Department of Biostatistics and Bioinformatics, Duke University, 2424 Erwin Road, Hock Plaza, Suite 1102, DUMC Box 2721, Durham, NC 27710, USA, jung0005@mc.duke.edu.

1994), those in biostatistics field have been sparse. Among some biostatistical examples are Jeong and Jung (2008) on two-sample comparison of median residual lifetime, and Jung, Jeong and Bandos (2009) extending the two-sample problem to regression analysis. More recently, residual lifetime analysis has been very popularly used in clinical trials to analyze the change in survival distribution of patients as a progress of disease (e.g. Zamboni et al. 2010; Zabor et al. 2013; Bischof et al. 2015; Mertens et al. 2015) under the name of conditional survival analysis. Let $T$ be the survival variable of a population with survivor function $S(t) = P(T \geq t)$. The $t$-year conditional survival distribution for patients who have survived for $t_0$ years, $P(T \geq t + t_0 | T \geq t_0)$, is denoted as $S(t|t_0) = S(t + t_0)/S(t_0)$ for $t \geq 0$. In the clinical literatures (e.g. Zabor et al. 2013; Mertens et al. 2015), investigators estimate the conditional survival distributions by replacing the survivor functions with their Kaplan-Meier (1958) estimates, but they do not provide a formal statistical testing to compare them between patient groups.

This paper can be regarded as a review paper supporting the analysis methods that are popularly used in medical field without theoretical justification. In this paper, we present analysis methods for conditional survival distributions including confidence interval of conditional survivor function for 1-sample problem, the conditional log-rank test for 2-sample test, and the conditional Cox (1972) proportional hazards model for regression analysis. We present simulation results to evaluate the performance of these methods. The proposed methods are demonstrated with real clinical data.

## 2 Analysis of Conditional Survival Distributions

### 2.1 One-Sample Problem

Suppose the lifetimes from n patients $T_1, \ldots, T_n$ are independent and identically distributed with survivor function $S(t) = P(T_i \geq t)$ and cumulative hazard function $\Lambda(t) = -\log S(t)$. From patient $i$, we observe $(X_i, \delta_i)$, where $X_i$ is the minimum of $T_i$ and censoring time $C_i$, and $\delta i$ is an event indicator taking 1 if patient $i$ had an event and 0 otherwise. We assume that censoring times are independent of the survival times.

For patients who have survived for at least $t_0 (\geq 0)$ years, the probability that they live additional t years, $S(t|t_0) = P(T_i \geq t + t_0 | T_i \geq t_0)$, is given as

$$S(t|t_0) = S(t + t_0)/S(t_0) \text{ for } t \geq 0 \quad (1)$$

by the definition of conditional probabilities. $S(t|t_0)$ is called the conditional survivor function for patients who have survived for $t_0$ years.

The conditional cumulative hazard function $\Lambda(t|t_0) = -\log S(t|t_0)$ is given as $\Lambda(t|t_0) = \Lambda(t + t_0) - \Lambda(t_0)$ from (1). Hence, for $t \geq 0$, the conditional hazard function $\lambda(t|t_0) = \partial\Lambda(t|t_0)/\partial t$ is identical to the unconditional (or marginal) hazard function $\lambda(t + t_0) = \partial\Lambda(t + t_0)/\partial t$. Jeong, Jung and Costantino (2008) propose a nonparametric inference method on conditional median residual lifetime $\theta$ that satisfies $S(\theta|t_0) = 1/2$.

Let $\hat{S}(t)$ denote the Kaplan-Meier estimator of $S(t)$. Then, $S(t|t_0)$ is consistently estimated by

$$\hat{S}(t|t_0) = \hat{S}(t + t_0)/\hat{S}(t_0) \text{ for } t \geq 0.$$

From Corollary 3.2.1 of Fleming and Harrington (1991), we have

$$\sqrt{n}\begin{bmatrix} \hat{S}(t + t_0) - S(t + t_0) \\ \hat{S}(t_0) - S(t_0) \end{bmatrix} = -\sqrt{n}\begin{bmatrix} S(t + t_0)\int_0^{t+t_0} Y(s)^{-1} dM(s) \\ S(t_0)\int_0^{t_0} Y(s)^{-1} dM(s) \end{bmatrix} + o_p(1) \quad (2)$$

where $M(t) = \int_0^t \{dN(s) - Y(s)d\Lambda(s)\}$ is a 0-mean martingale, and $N(t) = \sum_{i=1}^n \delta_i I(X_i \leq t)$ and $Y(t) = \sum_{i=1}^n I(X_i \geq t)$ denote the event and the at-risk processes, respectively. Let $y(t)$ denote the uniform limit of $n^{-1}Y(t)$ for $t \leq \tau$, where $\tau$ denotes the minimum of the upper limits of the supports of censoring and survival distributions. By the martingale central limit theorem, $\sqrt{n}\{\hat{S}(t + t_0) - S(t + t_0), \hat{S}(t_0) - S(t_0)\}$ converges to $N(0, \Sigma)$ in distribution, where

$$\Sigma = \begin{bmatrix} S(t + t_0)^2 \int_0^{t+t_0} y(s)^{-1} d\Lambda(s) & S(t_0)S(t + t_0)\int_0^{t_0} y(s)^{-1} d\Lambda(s) \\ S(t_0)S(t + t_0)\int_0^{t_0} y(s)^{-1} d\Lambda(s) & S(t_0)^2 \int_0^{t_0} y(s)^{-1} d\Lambda(s) \end{bmatrix},$$

which can be consistently estimated by replacing $y(t)$, $S(t)$ and $d\Lambda(t)$ with their consistent estimators $Y(t)/n$, $\hat{S}(t)$ and $Y(t)^{-1}dN(t)$, respectively. That is, a consistent estimator of $\Sigma$ is given as

$$\hat{\Sigma} = n\begin{bmatrix} \hat{S}(t + t_0)^2 \hat{a}(t + t_0) & \hat{S}(t_0)\hat{S}(t + t_0)\hat{a}(t_0) \\ \hat{S}(t_0)\hat{S}(t + t_0)\hat{a}(t_0) & \hat{S}(t_0)^2 \hat{a}(t_0) \end{bmatrix},$$

where $n\hat{a}(t)$ is a consistent estimator of $\int_0^t y(s)^{-1} d\Lambda(s)$ that is given as

$$\hat{a}(t) = \int_0^t Y(s)^{-1} d\hat{\Lambda}(s) = \sum_{i=1}^n \frac{\delta_i I(0 \leq X_i \leq t)}{\left\{\sum_{j=1}^n I(X_j \geq X_i)\right\}^2}$$

by using the Nelson-Aalen (Nelson, 1969; Aalen, 1978) estimator $\hat{\Lambda}(s) = \int_0^t Y(s)^{-1} dN(s)$

The partial differentiation of $\theta_1/\theta_2$ with $\theta_1 = S(t + t_0)$ and $\theta_2 = S(t_0)$ is given as

$$\nabla = \left[\frac{1}{S(t_0)}, -\frac{S(t+t_0)}{S(t_0)^2}\right]^T.$$

Hence, by the delta-method, $\sqrt{n}\{\hat{S}(t|t_0) - S(t|t_0)\}$ converges to $N(0,\sigma^2)$ in distribution, where

$$\sigma^2 = \nabla^T \Sigma \nabla = \{S(t|t_0)\}^2 \int_{t_0}^{t+t_0} {}_{y(s)}{}^{-1} d\Lambda(s).$$

A consistent estimator $\hat{\sigma}^2$ can be estimated by replacing $S(t)$, $y(t)$ and $d\Lambda(t)$ with their consistent estimators $\hat{S}(t)$, $n^{-1}Y(t)$ and $Y(t)^{-1}dN(t)$, respectively. That is,

$$\hat{\sigma}^2 = n\hat{S}(t|t_0)^2 \sum_{i=1}^{n} \frac{\delta_i I(t_0 \le X_i \le t+t_0)}{\{\sum_{j=1}^{n} I(X_j \ge X_i)\}^2}.$$

A 100(1 — α)% confidence interval for the conditional survival probability $S(t|t_0)$ can be calculated using this asymptotic result, i.e.

$$\hat{S}(t+t_0)/\hat{S}(t_0) \pm z_{1-\alpha/2}\hat{\sigma},$$

where $z_{1-\alpha}$ denote the 100(1 — α) percentile of the standard normal distribution.

These inferences are available only when there are patients who are at risk at t0 in the data set. In fact, these inferences can be much simplified. Suppose that $P(T_i \ge t_0)P(C \ge t_0) > 0$, i.e. the maximum survival time is longer than $t_0$ and some patients are followed for longer than $t_0$. Then, for $t > 0$, we have

$$P(T_i \ge t+t_0 | X_i \ge t_0) = \frac{P(T_i \ge t+t_0, X_i \ge t_0)}{P(X_i \ge t_0)}. \quad (3)$$

The right hand side of (3) equals

$$\frac{P(T_i \ge t+t_0, T_i \ge t_0, C_i \ge t_0)}{P(T_i \ge t_0, C_i \ge t_0)} = \frac{P(T_i \ge t+t_0, C_i \ge t_0)}{P(T_i \ge t_0, C_i \ge t_0)} = \frac{P(T_i \ge t+t_0)}{P(T_i \ge t_0)}$$

since $X_i = \min(T_i, C_i)$, and $T_i$ and $C_i$ are independent. Hence, we have $S(t|t_0) = P(T_i \ge t+t_0 | X_i \ge t_0)$, denoting the survival probability at $t+t_0$ for patients who are at risk at $t_0$. This implies that the conditional survival probability $S(t|t_0)$ can be estimated by calculating the Kaplan-Meier estimator at $t+t_0$ from the patients who are at risk at $t_0$.

This relationship becomes clearer by the definition of Kaplan-Meier estimator. For simplicity of notation, suppose that there are no ties among $X_1, \ldots, X_n$. Then, by the definition of Kaplan-Meier estimator, (2) is expressed as

$$\hat{S}(t|t_0) = \frac{\prod_{i:X_i \le t+t_0}\{1-\delta_i/Y(X_i)\}}{\prod_{i:X_i \le t_0}\{1-\delta_i/Y(X_i)\}} = \prod_{i:t_0 < X_i \le t+t_0}\{1-\delta_i/Y(X_i)\},$$

which is the Kaplan-Meier estimator at $t + t_0$ calculated from the data set consisting of patients who are at risk at time $t_0$, and its variance is consistently estimated by $n^{-1}\hat{\sigma}^2$. Hence, one-sample inference of conditional survival distribution using the delta-method will be identical to that based on the standard (or, unconditional) survival distribution using Kaplan-Meier estimator to the subset of data consisting of the patients who are at risk at t0. These results hold with tied survival data too.

An alternative confidence interval for $S(t|t_0)$ can be obtained using the Fieller's (1954) method. Let $\rho = \hat{S}(t_0)/\hat{S}(t+t_0)$. By using the asymptotic result for Kaplan-Meier estimator, $\sqrt{n}\{\rho\hat{S}(t+t_0) - \hat{S}(t_0)\}$ is asymptotically normal with mean 0, and its variance can be consistently estimated by $n(\rho^2\hat{\sigma}_{22} - 2\rho\hat{\sigma}_{12} + \sigma_{11})$, where $\hat{\sigma}_{ij}$ is the $(i, j)$-component of $n^{-1}\hat{\Sigma}$. Hence, we have

$$P\left(-z_{1-\alpha/2} < \frac{\rho\hat{S}(t+t_0) - \hat{S}(t_0)}{\sqrt{\rho\sigma_{11} - 2\rho\sigma_{12} + \sigma_{22}}} < z_{1-\alpha/2}\right) = 1-\alpha. \quad (4)$$

We can obtain a $100(1-\alpha)\%$ confidence interval of $S(t|t_0)$ by solving the equation within the probability of (4) with respect to $p$, i.e.

$$\frac{f_1 \pm \sqrt{f_1^2 - f_0 f_2}}{f_2},$$

where $f_0 = \hat{S}(t+t_0)^2 - \hat{\sigma}_{11}z^2_{1-\alpha/2}$, $f_1 = \hat{S}(t+t_0)\hat{S}(t_0) - \hat{\sigma}_{12}z^2_{1-\alpha/2}$ and $f_2 = \hat{S}(t_0)^2 - \hat{\sigma}_{22}z^2_{1-\alpha/2}$. This formula gives an appropriate confidence interval when $f_1^2 > f_0 f_2$

## 2.2 Two-Sample Log-Rank Test

Suppose that $n_k$ patients are randomized to arm $k(= 1, 2)$, and the survival time $T_{k1}, \ldots, T_{kn_k}$ from the $n_k$ patients of arm $k$ are independent and identically distributed with survivor function $Sk(t) = P(T_{ki} \ge t)$ and cumulative hazard function $\Lambda_k(t) = -\log S_k(t)$. From patient

$i(= 1,\text{–}, n_k)$ in arm $k(= 1, 2)$, we observe $(X_{ki}, \delta_{ki})$, where $X_{ki}$ is the minimum of $T_{ki}$ and censoring time $C_{ki}$, and $\delta_{ki}$ is an event indicator taking 1 if the patient had an event and 0 otherwise. We assume that the censoring times are independent of the survival times within each arm. Let $N_k(t) = \sum_{i=1}^{n_k} \delta_{ki} I(X_{ki} \le t)$ and $Y_k(t) = \sum_{i=1}^{n_k} I(X_{ki} \ge t)$ denote the event and the at-risk processes for arm $k$, respectively. Also, let $n = n_1 + n_2$, $N(t) = N_1(t) + N_2(t)$ and $Y(t) = Y_1(t) + Y_2(t)$.

For conditional survivor function $S_k(t|t_0) = S_k(t + t_0)/S_k(t_0)$, we want to derive a log-rank test to test $H_0 : S_1(t|t_0) = S_2(t|t_0)$ for all $t \ge 0$ against $H_1 : S_1(t|t_0) \ne S_2(t|t_0)$ for some $t \ge 0$. From the previous section, $d\Lambda(t|t_0) = d\Lambda(t + t_0)$ for $t \ge 0$, so that we can consider a log-rank test for comparing conditional survival distributions for patients who have survived over $t_0$,

$$W_{t_0} = \sqrt{n} \int_0^\infty H(t + t_0)\{d\hat{\Lambda}_1(t|t_0) - d\hat{\Lambda}_2(t|t_0)\} = \sqrt{n} \int_0^\infty H(t + t_0)\{d\hat{\Lambda}_1(t + t_0) - d\hat{\Lambda}_2(t + t_0)\},$$

which is identical to

$$W_{t_0} = \sqrt{n} \int_{t_0}^\infty H(t)\{d\hat{\Lambda}_1(t) - d\hat{\Lambda}_2(t)\}$$

where $H(t)$ is a predictable function that is uniformly convergent to $h(t)$ over $[t_0, \tau]$ and $\tau$ is the minimum of the supports of the censoring and survival distributions. The logrank statistic (Peto and Peto, 1972) uses $H(t) = n^{-1} Y_1(t) Y_2(t)/Y(t)$, the Gehan-Wilcoxon test (Gehan, 1965) uses $H(t) = n^{-2} Y_1(t) Y_2(t)$, and the Prentice-Wilcoxon test (Prentice, 1978) uses $H(t) = n^{-1}\hat{S}^-(t) Y_1(t) Y_2(t)/Y(t)$, where $\hat{S}^-$ is the left-continuous version of the Kaplan-Meier (1958) estimate from the pooled data.

Note that $W_{t_0}$ has the same expression as the standard rank tests $W_0$ except that the range of the integration is restricted to $[t_0, \infty)$. Using the same arguments as those used for the standard rank tests (e.g. Gill, 1980; Fleming and Harrington, 1991), we can show that $W_{t_0}/\hat{\sigma}_{t_0}$ is asymptotically standard normal with

$$\hat{\sigma}^2_{t_0} = n \int_{t_0}^\infty \frac{H(t)^2}{Y_1(t) Y_2(t)} dN(t)$$

under $H_0$. Hence, we reject $H_0$ in favor of $H_1$, if $\left|W_{t_0}/\hat{\sigma}_{t_0}\right| > z_{1-\alpha/2}$ with two-sided type I error rate $\alpha$.

For example, for the conditional log-rank test, we have

$$W_{t_0} = \frac{1}{\sqrt{n}} \int_{t_0}^{\infty} \frac{Y_1(t) Y_2(t)}{Y(t)} \left\{ d\,\widehat{\Lambda}_1(t) - d\,\widehat{\Lambda}_2(t) \right\}$$

$$= \frac{1}{\sqrt{n}} \left\{ \sum_{i=1}^{n_1} \delta_{1i} I\left(X_{1i} \geq t_0\right) \frac{\sum_{i'=1}^{n_2} I\left(X_{2i'} \geq X_{1i}\right)}{\sum_{k=1}^{2} \sum_{i'=1}^{n_k} I\left(X_{ki'} \geq X_{1i}\right)} - \sum_{i=1}^{n_2} \delta_{2i} I\left(X_{2i} \geq t_0\right) \frac{\sum_{i'=1}^{n_1} I\left(X_{1i'} \geq X_{2i}\right)}{\sum_{k=1}^{2} \sum_{i'=1}^{n_k} I\left(X_{ki'} \geq X_{2i}\right)} \right\}$$

and

$$\hat{\sigma}^2_{t_0} = \frac{1}{n} \int_{t_0}^{\infty} \frac{Y_1(t) Y_2(t)}{Y(t)^2} dN(t)$$

$$= \frac{1}{n} \sum_{k=1}^{2} \sum_{i=1}^{n_k} \delta_{ki} I\left(X_{ki} \geq t_0\right) \frac{\left\{ \sum_{i'=1}^{n_1} I\left(X_{1i'} \geq X_{ki}\right) \right\} \left\{ \sum_{i'=1}^{n_2} I\left(X_{2i'} \geq X_{ki}\right) \right\}}{\left\{ \sum_{k'=1}^{2} \sum_{i'=1}^{n_{k'}} I\left(X_{k'i'} \geq X_{ki}\right) \right\}^2}.$$

From the expression of $W_{t_0}$ and $\hat{\sigma}^2_{t_0}$, it is obvious that the conditional log-rank test at $t_0$ can be carried out by applying the standard log-rank test to the data set consisting of patients who are at risk at $t_0$, $\mathcal{D}(t_0) = \left\{ \left(X_{ki}, \delta_{ki}\right) : X_{ki} \geq t_0, k = 1, 2, i = , \ldots, n_k \right\}$. The 2-sample conditional log-rank test can be easily extended to the log-rank test for $K$-sample cases with $K > 2$. These results holds for other types of conditional rank tests.

### 2.3 Regression Method

From patient $i = 1, \ldots, n$, we observe covariates $z_i = (z_{1i}, \ldots, z_{mi})^T$ together with the minimum of the survival and censoring times $X_i$ and event indicator $\delta_i$. We assume that, given $z_i$, the survival and censoring times are independent. Suppose that the conditional survival distribution for patients who have survived over $t_0$ has a proportional hazards model

$$\lambda_i\left(t \middle| t_0\right) = \lambda_0\left(t \middle| t_0\right) \exp\left(\beta^T z_i\right) \quad (5)$$

for $t \geq 0$, where $\lambda_i(t|t_0)$ denotes the baseline conditional hazard function. As was shown in the previous sections, the conditional hazard function $\lambda_i(t|t_0)$ is identical to the unconditional hazard function $\lambda_i(t + t_0)$, so that (5) can be expressed as the regular proportional hazards model

$$\lambda_i(t) = \lambda_0(t) \exp\left(\beta^T z_i\right)$$

for $t \geq t_0$. Hence, if the (unconditional) survival distribution has a proportional hazards model with constant covariate effect over the whole time span, then the conditional survival distribution for any $t_0 (> 0)$ has the same proportional hazards model. However, if the covariate effect changes over time, then the regression model for conditional survival distribution changes in $t_0$.

The partial score and information matrix (Cox 1972) are given as

$$U_{t_0}(\beta) = \sum_{i=1}^{n} \int_{t_0}^{\infty} \left( Z_i - \frac{\sum_{j=1}^{n} Z_j Y_j(t) e^{\beta^T Z_j}}{\sum_{j=1}^{n} Y_j(t) e^{\beta^T Z_j}} \right) Y_i(t) dN_i(t)$$

and

$$A_{t_0}(\beta) = \sum_{i=1}^{n} \int_{t_0}^{\infty} \left\{ \frac{\sum_{j=1}^{n} Z_j^{\otimes 2} Y_j(t) e^{\beta^T Z_j}}{\sum_{j=1}^{n} Y_j(t) e^{\beta^T Z_j}} - \left( \frac{\sum_{j=1}^{n} Z_j Y_j(t) e^{\beta^T Z_j}}{\sum_{j=1}^{n} Y_j(t) e^{\beta^T Z_j}} \right)^{\otimes 2} \right\} Y_i(t) dN_i(t)$$

respectively, where $z^{\otimes 2} = zz^{\mathrm{T}}$ for a column vector $z$. Using the same asymptotic theory for the Cox regression method, we can show that $\sqrt{n}\left(\hat{\beta} - \beta\right)$ is approximately normal with mean 0 and its variance-covariance matrix can be consistently approximated by $nA_{t_0}^{-1}\left(\hat{\beta}\right)$.

For a univariate proportional hazards model with a dichotomous covariate, it is easy to show that the partial score test under $\beta = 0$, $U_{t_0}(0)/\sqrt{A_{t_0}(0)}$, is identical to the conditional log-rank test, $W_{t_0}/\hat{\sigma}_{t_0}$, that was discussed in the previous section. From the expression of the partial score and information, it is obvious that the conditional Cox regression model (5) can be fitted by applying the standard Cox regression method to the data set consisting of the patients who are at risk at t0, i.e. $D(t_0) = \{(X_i, \delta_i, z_i) : X_i \geq t_0, i = 1, \ldots, n\}$. Kurta et al. (2014) proposed this analysis method without any theoretical justification.

## 3 Numerical Studies

### 3.1 Simulations

We want to show that the standard inference methods using the subset of data appropriately reflect the conditional survival distribution concept. At first, we conduct simulations on one-sample problems using a piecewise exponential distribution with survivor function

$$S(t) = \begin{cases} \exp\left(-\lambda_1 t\right) & \text{if } 0 \leq t < 2 \\ \exp\left\{-\lambda_2 t - 2\left(\lambda_1 - \lambda_2\right)\right\} & \text{if } t \geq 2 \end{cases}$$

This distribution has a hazard function of $\lambda_1$ for $0 \leq t \leq 2$ and $\lambda_2$ for $t \geq 2$. By choosing $\lambda_1 = 0.3466$ and $\lambda_2 = \lambda_1/2$, the median survival for the whole patients is 2 years at the baseline, while that for those who have survived the first $t_0 = 2$ years is 4 years starting from the 2-year time point. Survival times are generated from $S(t)$ and censoring times from $U(0,a)$, where $a$ is chosen for 30% of censoring rate. With $a$ fixed at this value, we consider censoring distribution $U(b, a + b)$ with $b$ chosen for 15% of censoring. We generate 10,000

simulation samples of size $n = 200$. From each sample, we estimate $S(t|t_0)$ and its 95% confidence interval by the delta-method and Fieller's method for $t_0 = 1, 2, 3$, and 4, and $t = 1, 2, \ldots, 5 - t_0$.

Table 1 reports mean bias and the sample standard deviation over the simulation samples (SSD) of the estimator, $\hat{S}(t|t_0)$, and empirical coverage probabilities of the two confidence interval methods. Table 1 also reports the mean of the standard deviation (MSD) of $\hat{S}(t|t_0)$ estimated by the delta method over the simulation samples. We observe that the estimated conditional probabilities have very small bias and the bias tends to increase in censoring proportion. As expected, SSD and MSD tend to increase in censoring proportion. They are very close for each simulation setting, but MSD is slightly smaller than SSD. This may result in slightly anti-conservative empirical coverage probability of the confidence intervals by the delta method. As $t_0$ and $t$ increase, the number of subjects at risk decreases and the bias tends to be negative. The two confidence interval methods have empirical coverage probabilities close to the nominal 95% overall, but Fieller's method always have slightly larger average length and more accurate coverage probability than the delta method. It is known that Fieller's method usually provides better large sample approximation for a ratio of parameters than the delta method, e.g. Herson (1975). For each method, the average length increases in censoring proportion.

Now, we investigate the finite sample properties of the two-sample log-rank test on conditional distributions. Suppose that arm 1 has an exponential distribution with $S_i(t) = \exp(-\lambda_1 t)$ for $t \geq 0$, and arm 2 has a piecewise exponential distribution with survivor function

$$S_2(t) = \begin{cases} \exp(-\lambda_2 t) & \text{if } 0 \leq t < 2 \\ \exp\{-\lambda_1 t - 2(\lambda_2 - \lambda_1)\} & \text{if } t \geq 2 \end{cases}$$

Note that, if $\lambda_1 \neq \lambda_2$, then the two arms have different survival distributions, but with $t_0 \geq 2$, their conditional distributions are identical with $S_k(t|t_0) = \exp(-\lambda_1 t)$ and $\lambda_k(t|t_0) = \lambda_k(t + t_0) = \lambda_1$ for $t \geq 0$. So, the log-rank test will have some power for $t_0 < 2$, but not for $t_0 \geq 2$. We set $\lambda_1 = 0.3466$, $\lambda_2 = \lambda_1/2$, and $n_1 = n_2 = 100$, 150 or 200. We consider 15% and 30% censoring by uniform censoring variables as in the previous simulations. We generate 10,000 samples, apply the 2-sample log-rank test with 2-sided $\alpha = 0.05$ for $t_0 = 0,1, 2, 3, 4$ to each sample, and estimate the empirical power as the proportion of samples that the log-rank test rejects the null hypothesis that two arms have the same conditional survival distributions. Note that the test with $t_0 = 0$ corresponds to the standard log-rank test to compare two unconditional distributions.

Table 2 reports the empirical power of the log-rank tests. As expected, the empirical power of the conditional log-rank test is close to the nominal level $\alpha = 0.05$ with $t_0 \geq 2$, for which two conditional distributions are identical. However, with $t_0 = 0$ and 1, it has some power, and the power becomes higher with a smaller $t_0(= 0)$ since the time interval over which the two conditional distributions are different is wider in this case. The empirical power for $t_0 =$

0 and 1 also increases in $n(= n_1 + n_2)$, while that for $t_0 \geq 2$ is close to the nominal $\alpha = 5\%$ regardless of the sample size. We observe that the power with $t_0 < 2$ does not much depend on the censoring proportion under the simulation setting.

We consider two regression models for simulations on Cox regression analysis of conditional survival distributions. In Model 1, given covariate value $z_i$, the hazard function is given as

$$\lambda_i(t) = \begin{cases} \lambda_0 \exp(\beta z_i) & \text{if } 0 \leq t < 2 \\ \lambda_0 & \text{if } t \geq 2 \end{cases}$$

Since the hazard function for $t \geq 2$ does not depend on $z_i$, the conditional Cox regression with $t_0 > 2$ will be free of the covariate. The cumulative hazard function is given as

$$\Lambda_i(t) = \begin{cases} \lambda_0 t \exp(\beta z_i) & \text{if } 0 \leq t < 2 \\ 2\lambda_0 \exp(\beta z_i) + \lambda_0 (t-2) & \text{if } t \geq 2 \end{cases}$$

and the survivor function is given as

$$S_i(t) = e^{-\Lambda_i(t)} = \begin{cases} \exp\{-\lambda_0 t \exp(\beta z_i)\} & \text{if } 0 \leq t < 2 \\ \exp\{-2\lambda_0 \exp(\beta z_i) - \lambda_0 (t-2)\} & \text{if } t \geq 2 \end{cases}$$

For a $U(0,1)$ random variable $U_i$, we generate $T_i$ by solving $S_i(T_i) = U_i$. We set $\beta = 0.3$ and $\lambda_0 = 0.3$ for Model 1.

In Model 2, we consider a piecewise exponential distribution with a time-dependent covariate effect: for patient i with covariate value $z_i$, $\lambda_i(0) = 0$,

$$\lambda_i(t) = \lambda_0 \exp(\beta / j z_i) \quad \text{if } j - 1 < t \leq j$$

for $j = 1, 2, \ldots$ For this model, the covariate effect $\beta/j$ decreases in $t$. The cumulative hazard function given $z_i$ is,

$$\Lambda_i(t) = \lambda_0 \sum_{j=1}^{k} \exp(\beta / j z_i) - (k - t) \lambda_0 \exp(\beta / k z_i) \quad \text{if } k - 1 < t \leq k.$$

Since $\Lambda_i(t) = -\log S_i(t)$ and $S_i(T_i) \sim U(0,1)$ for $T_i$ with cumulative survivor function $S_i(t)$, we generate $T_i$ by solving equation $\Lambda_i(T_i) = -\log U_i$ for $U_i \sim U(0,1)$, i.e.

$$T_i = k - \frac{\lambda_0 \sum_{j=1}^{k} \exp(\beta / j z_i) + \log U_i}{\lambda_0 \exp(\beta / k z_i)} \quad \text{if } \lambda_0 \sum_{j=1}^{k-1} \exp(\beta / j z_i) < -\log U_i \leq \lambda_0 \sum_{j=1}^{k} \exp(\beta / j z_i).$$

We set $\lambda_0 = 0.3$ and $\beta = 0.4$ for Model 2.

For each of the survival models, we generate 10,000 simulation samples of size n = 500 and generate 15% and 30% censoring from uniform distributions $U(b, a + b)$ as in the previous simulations. For each subject, covariate $z_i$ is generated from the standard normal distribution. From each sample, conditioning on ($T_i \geq t_0$) with $t_0$ = 0, 1, 2, 3, or 4, we fit a proportional hazards model with a time-independent regression coefficient, estimate the regression coefficient, and test on $H_0 : \beta = 0$ with 2-sided $\alpha = 0.05$.

Table 3 report the mean regression estimate and empirical power under the two models. For Model 1, with $t_0 \geq 2$, the mean regression estimate is close to 0 and the empirical power is close to the nominal 0.05 level as expected. But, for $t_0 < 2$, the regression estimate is smaller than $\beta = 0.3$ since the covariate effect is diluted over the time interval $t \geq 2$ which has no covariate effect. The regression estimate is smaller with $t_0 = 1$ than with $t_0 = 0$ since the former case has a narrower time interval with non-zero regression coefficient. With 30% of censoring, the regression estimate is larger since the additional censoring over 15% censoring occurs over $t \geq 2$ ($b = 2.2$ for Model 1) for which the covariate has no effect. For Model 2, we observe that the regression estimate decays in $t_0$. And the decaying trend is more prominent with 30% censoring since the additional censoring occurs after $b = 2.1$ where the covariate effect is smaller than that over the earlier time interval. The empirical power quickly decreases in $t_0$ since both the mean covariate effect and the number of observations used in analysis decrease. However, the decrease of power in censoring proportion is smaller since the additional censoring occurs over the time interval with smaller covariate effect.

### 3.2 Real Data Analysis

Kim et al. (2016) report analysis results of a retrospective record study on 723 lung adeno-carcinoma patients. All the patients underwent complete resection and mediastinal lymph node dissection with or without postsurgical adjuvant therapy. From each patient, overall survival (OS), time to death of any cause from surgery for tumor resection, and progression-free survival (PFS), time to tumor progression, were observed as outcomes together with risk factors including ECOG performance score (PS), with or without adjuvant chemotherapy (adj), tumor-shadow disappearance ratio (TDR) on CT value, and maximum standardized uptake value (SUV) on 18F-uoro-2-deoxyglucose (FDG)-PET/CT. SUV is log-transformed to lower the effect of outliers. The objectives of the study is to associate OS and PFS with the latter four clinical and image predictors using conditional survival analysis. We report the analysis results on OS to illustrate the conditional survival analysis methods. At first, patients are partitioned into two groups by PS = 0 and PS ≥ 1. Figure 1 displays the conditional survivor functions of the two PS groups and conditional log-rank p-value for t0 = 0,1,…, 6. We observe that the effect of PS decays as time passes from surgery and becomes insignificant for patients who have survived for $t_0 = 4$ years or longer. For each $t_0$ value, we regress the conditional survival at $t_0 = 0, 1, \ldots, 6$ on these four covariates using a multivariate proportional hazards model. Figure 2 displays the regression estimate of each covariate and its 95% confidence interval against $t_0$. The covariate effect diminishes for longer survivors except history of adjuvant therapy which has a strong and consistent

negative effect on OS. A high log-SUV tends to be associated with shorter OS, but it is not so significant for the $t_0$ values considered. Poor PS is significantly associated with poor OS until $t_0$ = 2 years and its effect becomes weak after $t_0$ = 4 years. High TDR is associated with longer until about $t_0$ = 4 years, but its effect diminishes among survivors over $t_0$ = 5 years.

## 4 Conclusions

Conditional survival analysis has been popularly used to investigate the long term effect of treatment and baseline characteristics on the prognosis in clinical researches. As a reviewer points out, this analysis provides some new insight on the difference and effect of non-proportional hazards, early and late treatment and baseline patient characteristics. One meaningful scenario is that an aggressive surgical treatment may have a high early mortality, but leads to much higher survival or even cure after the treatment period, while a chemotherapy does not have a severe treatment-related mortality, but leads to a moderate treatment effect over a long time span. The conditional survival analysis would be particularly useful to compare this kind of early and late survival benefit between treatments.

Without any theoretical justification, investigators have applied the standard survival analysis methods, such as log-rank test and Cox regression, to the data removing the patients whose censoring or event times are shorter than $t_0$ claiming that this results in the survival distributions of the survivors over $t_0$. This paper is to theoretically justify this claim. We have reviewed inference methods of one-sample, two-sample and regression analysis for conditional survival distributions. For a reliable estimation of $S(t|t_0)$, we need enough number of patients who are at risk at $t_0$ and enough number of patients followed for at least $t + t_0$ unless all patients have events before this time point. Hence, a conditional survival analysis will not be available unless the follow-up period is long enough. Since the conditional survival analyses among the patients who survived over $t_0$ are identical to the standard survival analyses using the data set consisting of patients who are at risk at time $t_0$. Hence, we can conduct any conditional survival analysis using existing statistical softwares, such as SAS or SPSS, with the standard survival analysis procedures. These methods are based on large sample theory. Through simulations, we find that these methods accurately reflect the change in risk function over time and have good finite sample properties.

If the marginal survival distribution satisfies a proportional hazards model (PHM) assumption with time-fixed covariate effect, then the regression estimates from conditional survival analysis will give similar regression estimates for various $t_0$ values. We may be able to develop a goodness of fit test for PHM assumption of a marginal survival distribution using this concept. Jung and Wieand (1999) propose a goodness of fit test for PHM using a similar approach. By plotting the trend of regression estimates of conditional survival analysis over $t_0$, we can also model the time trend of covariates with time-varying effect, refer to Therneau and Grambsch (2000).

## Acknowledgments

Funding

This research was supported by a grant from the National Cancer Institute (CA142538–01).

## REFERENCES


Aalen OO (1978). Nonparametric estimation of partial transition probabilities in multiple decrement models. Annals of Statistics 6:534–545.

Bischof DA, Kim Y, Dodson R, Jimenez MC, Behman R, Cocieru A, Fisher SB,Groeschl RT, Squires MH, Maithel SK, Blazer DG, Kooby DA, Gamblin TC, Bauer TW, Quereshy FA, Karanicolas PJ, Law CH, Pawlik TM (2015). Conditional disease-free survival after surgical resection of gastrointestinal stromal tumors: A multi-institutional analysis of 502 patients. JAMA Surgery 150:299–306. [PubMed: 25671681]

Bryson C, Siddiqui MM (1969). Some criteria for aging. Journal of the American Statistical Association 64:1472–1483.

Cox DR (1972). Regression models and life tables. Journal of the Royal Statistical Society, Ser. B 34:187–220.

Fieller EC (1954). Some problems in interval estimation. Journal of the Royal Statistical Society, Ser. B bf 16:175185.

Fleming TR, Harrington DP (1991). Counting Processes and Survival Analysis. New York: Wiley.

Gehan EA (1965). A generalized Wilcoxon test for comparing arbitrarily single censored samples. Biometrika 52:203–223. [PubMed: 14341275]

Gill RD (1980). Censoring and Stochastic Integrals Mathematical Centre Tracts 124, Mathematisch Centrum, Amsterdam.

Herson J (1975). Fieller's theorem versus the delta method for significance intervals for ratios. Journal of Statistical Computing and Simulation 3:265–274.

Hollander E, Proschan F (1975). Tests for mean residual life. Biom etrika 62:585–593.

Jeong JH, Jung SH, Costantino JP (2008). Nonparametric inference on median residual life function. Biometrics 64:157–163. [PubMed: 17501936]

Jung SH, Jeong JH, Bandos H (2009). Regression on median residual life. Biometrics 65:1203–1212. [PubMed: 19432781]

Jung SH, Wieand S (1999). Analysis of Goodness-of-Fit for Cox Regression Model. Statistics and Probability Letters 41:379–82.

Kaplan EL, Meier P (1958). Nonparametric estimation from incomplete observations. Journal of American Statistical Association 53:457481.

Kim W, Lee HY, Jung SH, Kim HK, Choi YS, Kim J, Zo J, Shim YM, Han J, Jeong JY, Choi JY, Lee KS (2016). Dynamic prognostication using conditional survival analysis for patients with operable lung adenocarcinoma. Oncotarget DOI: 10.18632/oncotarget.12920.

Kurta ML, Edwards RP, Moysich KB, McDonough K, Bertolet M, Weissfeld JL, Catov JM, Modugno F, Bunker CH, Ness RB, Diergaarde B (2014). Prognosis and conditional disease-free survival among patients with ovarian cancer. Journal of Clinical Oncology 32:4102–4112. [PubMed: 25403208]

Mertens AC, Yong J, Dietz AC, Kreiter E, Yasui Y, Bleyer A, Armstrong GT, Robison LL, Wasilewski-Masker K (2015). Conditional survival in pediatric malignancies: Analysis of data from the childhood cancer survivor study and the surveillance, epidemiology, and end results program. Cancer 121:1108–1117. [PubMed: 25557134]

Muth EJ (1977). Reliability models with positive memory derived from the mean residual life function, In Theory and Applications of Reliability, (Edited by Tsokos CP and Shimi IN), PP 401–434, Academic Press.

Nelson W (1969). Hazard plotting for incomplete failure data. Journal of Quality Technology 1:27–52.

Peto R, Peto J (1972). Asymptotically efficient rank invariant test procedures (with discussion). Journal of the Royal Statistical Society, Ser. A 135:185–206.

Prentice RL (1978). Linear rank tests with right censored data. Biometrika 65:167–179.

Ruiz JM, Navarro J (1994). Characterization of distributions by relationships between failure rate and mean residual life. IEEE Transactions on Reliability 43:640–644.

Therneau TM, Grambsch PM (2000). Modeling Survival Data: Extending the Cox Model. Springer, New York, NY, USA.

Zabor EC, Gonen M, Chapman PB, Panageas KS (2013). Dynamic prognostication using conditional survival estimates. Cancer 119:3589–3592. [PubMed: 23913639]

Zamboni BA, Yothers G, Choi M, Fuller CD, Dignam JJ, Raich PC, Thomas CR, OConnell MJ, Wolmark N, Wang SJ (2010). Conditional survival and the choice of conditioning set for patients with colon cancer: An analysis of NSABP trials C-03 through C-07. Journal of Clinical Oncology 28:2544–2548. [PubMed: 20406942]

Zabor EC, Gonen M, Chapman PB, Panageas KS (2013). Dynamic prognostication using conditional survival estimates. Cancer 119:3589–3592. [PubMed: 23913639]

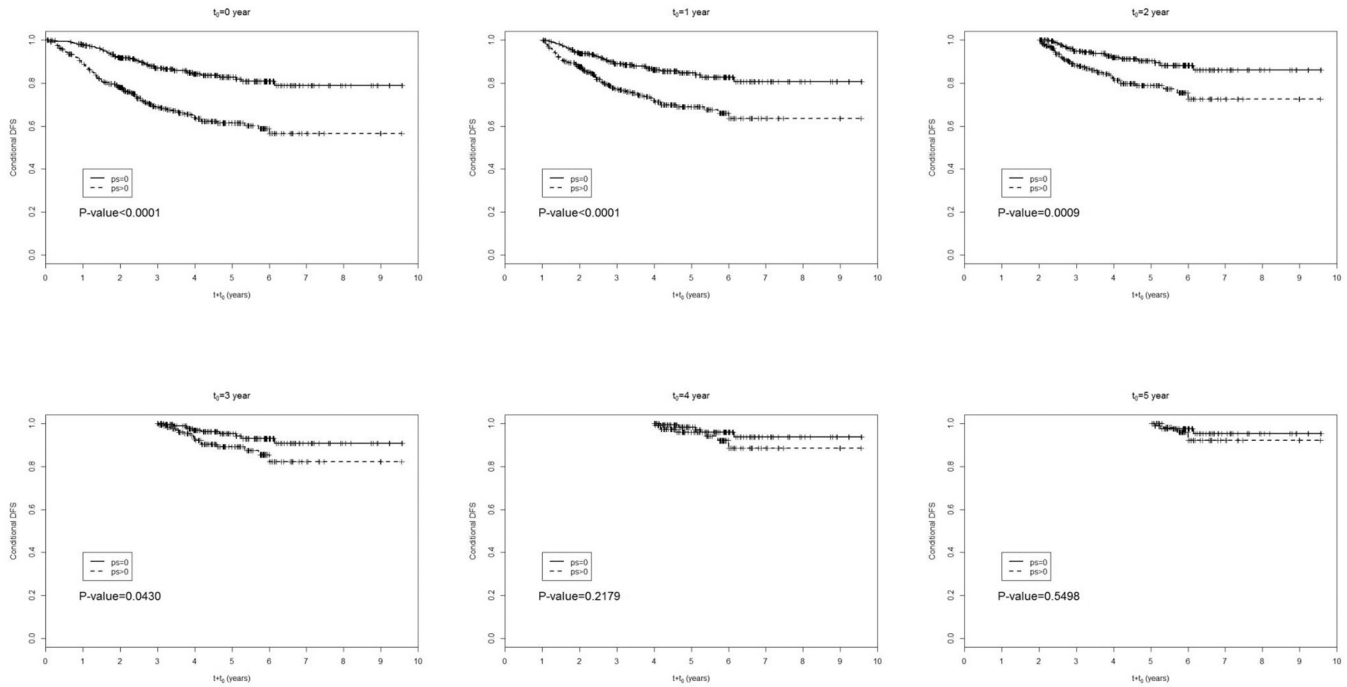


**Figure 1:**

Conditional survivor functions for $t_0$ = 0,1, 2, 3, 4, 5. The p-values are from the conditional log-rank test between PS=0 and PS> 0 groups

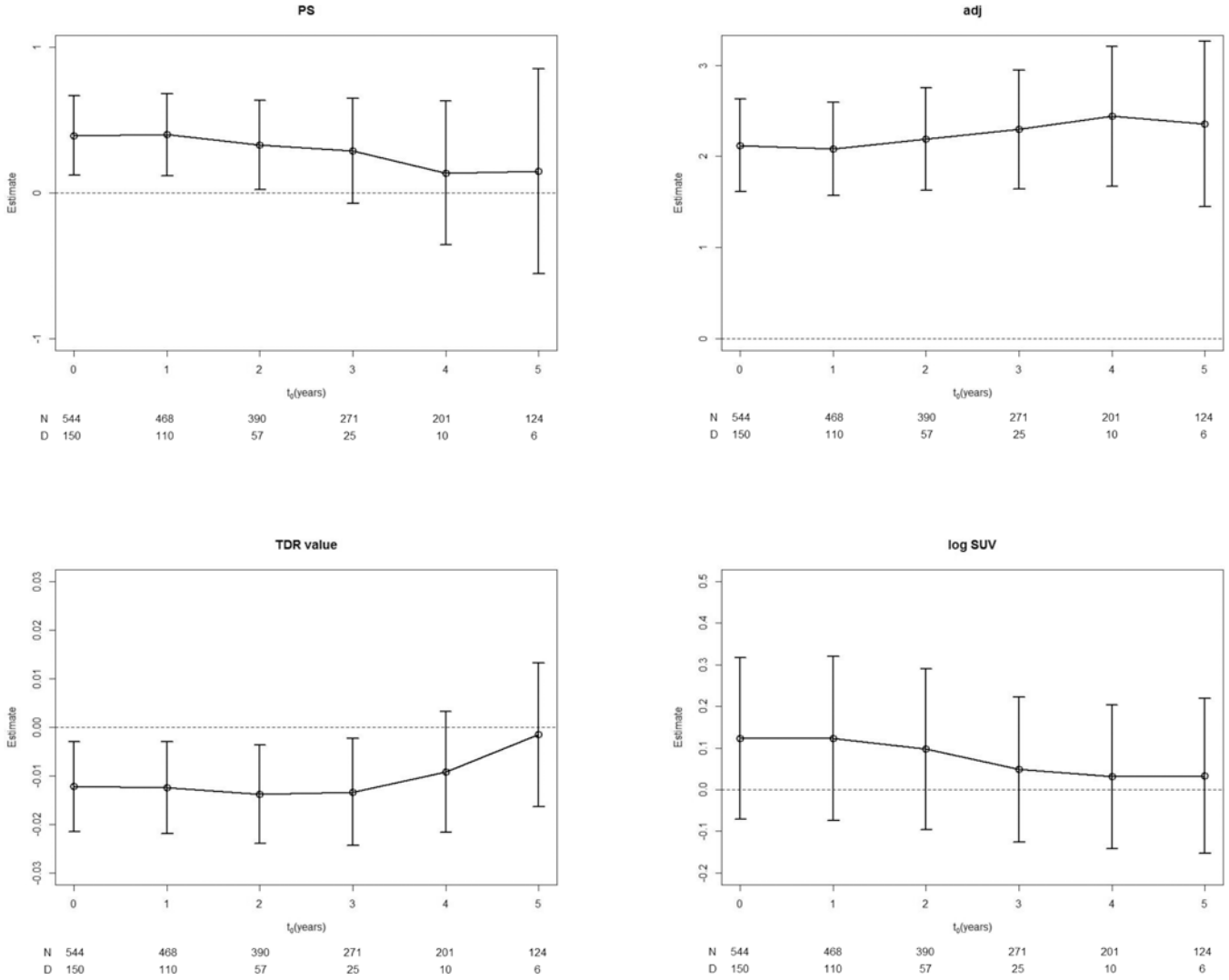


**Figure 2:**
Regression estimate of each covariate and its 95% confidence interval against $t_0$

**Table 1.**

Bias, SSD, and MSD of $\hat{S}(t|t_0)$, and average length and empirical coverage probability of 95% confidence intervals by Delta and Fieller methods

| | | | | | Delta Method | | Fieller Method | |
|---|---|---|---|---|---|---|---|---|
| $t_0$ | $t$ | Bias | SSD | MSD | ECP | Length | ECP | Length |
| Under 15% Censoring | | | | | | | | |
| 1 | 1 | 0.0004 | 0.0308 | 0.0299 | 0.9407 | 0.1096 | 0.9411 | 0.1098 |
| | 2 | 0.0005 | 0.0381 | 0.0379 | 0.9471 | 0.1486 | 0.9476 | 0.1489 |
| | 3 | 0.0001 | 0.0396 | 0.0390 | 0.9465 | 0.1528 | 0.9468 | 0.1530 |
| | 4 | 0.0000 | 0.0379 | 0.0371 | 0.9430 | 0.1453 | 0.9434 | 0.1456 |
| 2 | 1 | 0.0003 | 0.0386 | 0.0384 | 0.9470 | 0.1505 | 0.9477 | 0.1511 |
| | 2 | −0.0001 | 0.0440 | 0.0432 | 0.9442 | 0.1694 | 0.9451 | 0.1701 |
| | 3 | −0.0001 | 0.0434 | 0.0425 | 0.9433 | 0.1665 | 0.9435 | 0.1672 |
| 3 | 1 | −0.0004 | 0.0487 | 0.0475 | 0.9387 | 0.1860 | 0.9425 | 0.1879 |
| | 2 | −0.0004 | 0.0548 | 0.0536 | 0.9404 | 0.2099 | 0.9433 | 0.2120 |
| 4 | 1 | −0.0002 | 0.0607 | 0.0589 | 0.9350 | 0.2307 | 0.9399 | 0.2352 |
| Under 30% Censoring | | | | | | | | |
| 1 | 1 | 0.0003 | 0.0308 | 0.0299 | 0.9457 | 0.1170 | 0.9460 | 0.1172 |
| | 2 | 0.0006 | 0.0425 | 0.0416 | 0.9458 | 0.1631 | 0.9462 | 0.1634 |
| | 3 | 0.0000 | 0.0438 | 0.0432 | 0.9427 | 0.1696 | 0.9428 | 0.1699 |
| | 4 | −0.0001 | 0.0422 | 0.0415 | 0.9398 | 0.1629 | 0.9402 | 0.1632 |
| 2 | 1 | 0.0004 | 0.0434 | 0.0425 | 0.9438 | 0.1668 | 0.9451 | 0.1875 |
| | 2 | −0.0002 | 0.0488 | 0.0482 | 0.9420 | 0.1891 | 0.9433 | 0.1899 |
| | 3 | −0.0002 | 0.0486 | 0.0477 | 0.9398 | 0.1872 | 0.9406 | 0.1881 |
| 3 | 1 | −0.0006 | 0.0542 | 0.0531 | 0.9349 | 0.2085 | 0.9388 | 0.2109 |
| | 2 | −0.0007 | 0.0620 | 0.0604 | 0.9376 | 0.2369 | 0.9418 | 0.2397 |
| 4 | 1 | −0.0003 | 0.0686 | 0.0666 | 0.9320 | 0.2614 | 0.9374 | 0.2677 |

**Table 2.**

Empirical power of the conditional log-rank tests between $S_1(t)$ and $S_2(t)$ for $t_0 = 0,1, 2, 3, 4$

| $n$ | Censoring | $t_0 = 0$ | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|---|
| 200 | 15% | 0.5490 | 0.1971 | 0.0522 | 0.0510 | 0.0558 |
| | 30% | 0.5608 | 0.1853 | 0.0476 | 0.0549 | 0.0551 |
| 300 | 15% | 0.7202 | 0.2601 | 0.0508 | 0.0480 | 0.0485 |
| | 30% | 0.7324 | 0.2597 | 0.0515 | 0.0509 | 0.0516 |
| 400 | 15% | 0.8451 | 0.3345 | 0.0501 | 0.0535 | 0.0546 |
| | 30% | 0.8490 | 0.3206 | 0.0504 | 0.0502 | 0.0505 |

**Table 3.**

Mean regression estimate and empirical power of the conditional regression method

| | | Model 1 | | Model 2 | |
|---|---|---|---|---|---|
| **Censoring** | $t_0$ | **Mean Est** | **Power** | **Mean Est** | **Power** |
| 15% | 0 | 0.2022 | 0.8866 | 0.2610 | 0.9817 |
| | 1 | 0.1327 | 0.3706 | 0.1456 | 0.3912 |
| | 2 | 0.0036 | 0.0522 | 0.1076 | 0.1724 |
| | 3 | 0.0035 | 0.0530 | 0.0862 | 0.1043 |
| | 4 | 0.0026 | 0.0565 | 0.0717 | 0.0768 |
| 30% | 0 | 0.2191 | 0.8824 | 0.2787 | 0.9793 |
| | 1 | 0.1470 | 0.3538 | 0.1514 | 0.3409 |
| | 2 | 0.0021 | 0.0545 | 0.1102 | 0.1475 |
| | 3 | 0.0004 | 0.0560 | 0.0869 | 0.0951 |
| | 4 | −0.0033 | 0.0553 | 0.0692 | 0.0741 |