東海大學統計研究所

碩士論文

在 Partly Aalen 加成模型下兩累積風險函數之等效性檢定

Equivalence Tests for the difference of Two Cumulative Hazard Functions under Partly Aalen's Additive Hazard Model

指導教授:沈葆聖 博士

研究生:陳柏翰

中華民國一零八年一月

致謝

時間飛逝，轉瞬之間，兩年的研究所生活也即將劃下了句點，在即將畢業之時，驀然回首，內心充滿無限的惆悵，總覺得學問並沒有完全學完就即將畢業了，但是學海無涯，在漫長的人生中，畢業只是路上的中途站，總會尋到柳暗花明的那一刻。在惆悵之餘，內心也充滿無窮的感激。

首先，我打從心底感謝我的指導教授-沈葆聖老師的耐心教導，讓原本對存活分析研究懵懵懂懂的我有所進展，即使老師平常忙於教課與研究，但仍然把一些空餘時間分給我，細心的從頭慢慢的教導我，我能在今天順利從東海統計研究所畢業，沈老師絕對功不可沒。另外我也感謝兩位口試委員，林正祥教授和戴政教授，謝謝你們百忙之中能前來參加我的口試，並給予了我諸多關於論文上的建議，使我對這領域了解的更深入。

再來，在我兩年的研究所時間裡，我要感謝我的研究所同學和老師們，在課業上給予我許多幫助，還有感謝我的家人們，讓我能在經濟上無後顧之憂，能專心的在學習上。

Contents

Abstract

Clinical trials for determining equivalence of a new therapy with a standard therapy of proven efficacy have become increasingly important recently due to growing cost and ethical pressures to switch from an expensive and invasive standard therapy to a cheaper and less-invasive therapy. Statistical methods used for equivalence trial for survival response are often based on the method proposed by Wellek (1993) under proportional hazards (PH) model. Martinez et al. (2017) extended the result of Wellek (1993) to the case of proportional odds (PO) survival model. The Aalen's additive risk model has the feature that the influence of each covariate can vary separately and nonparametrically through time, which allows greater flexibility of temporal structure than PH and PO models. In this article, we propose equivalence tests for the difference of two cumulative hazard functions under partly Aalen's model, where the influence of some covariates varies nonparametrically over time, and that of the remaining covariates containing the indicator of treatment is constant. Simulation studies demonstrate that the proposed test performs well in practical situations.

Key Words: Semiparametric transformation model; maximum likelihood estimator; Type I error; critical region.

1 Introduction

In clinical trials, determining the equivalence or non-inferiority of a new drug (test drug) with an existing drug (reference drug) is an important topic since there are growing financial and ethical pressures to switch from an expensive and invasive standard therapy to a cheaper and less-invasive therapy. For the case of no covariates beyond treatment arms, let $S_0(t)$ and $S_1(t)$ denote the survival functions for standard treatment and new treatment. The aim of a clinical equivalence trial is usually to test the clinical scientific hypothesis

$$\tilde{H}_0 : |S_1(t) - S_0(t)| \geq \delta \ for\ some\ t \text{ versus } \tilde{H}_a : |S_1(t) - S_0(t)| < \delta \ for\ all\ t, \quad (1.1)$$

where δ is a specified cutoff value for equivalence between two survival functions $S_1(t)$ and $S_0(t)$. Under the assumption of proportional hazards (PH) model (Cox (1972)), i.e., $S_1(t) = (S_0(t))^{e^\beta}$, Wellek (1993) showed that the H_0 and H_a in (1.1) are equivalent to the statement

$$\tilde{H}_0^* : |\beta| \geq \log(1+\epsilon) \text{ versus } \tilde{H}_a^* : |\beta| < \log(1+\epsilon),$$

where β is the regression coefficient associated with the treatment group indicator, and ϵ is a simple strictly increasing function of δ and satisfies

$$\delta = e^{-\frac{1}{\epsilon}\log(1+\epsilon)} - e^{-\frac{1+\epsilon}{\epsilon}\log(1+\epsilon)}.$$

In many equivalence trials, the hazard functions of two treatment arms are not proportional over time. Martinez et al. (2017) demonstrated that the actual type I error rate for the procedure of Wellek (1993) is higher than the desired nominal rate when survival responses from two treatment arms satisfy the proportional odds (PO) model (Bennett (1983); Murphy et al. (1997)), i.e.,

$$S_0(t) = \frac{1}{1+R(t)} \quad \text{and} \quad S_1(t) = \frac{1}{1+R(t)e^\beta},$$

which implies

$$\frac{1-S_1(t)}{S_1(t)} = \theta\left[\frac{1-S_0(t)}{S_0(t)}\right],$$

where $\theta = e^\beta$ is the time-constant survival odds ratio between new treatment and standard treatment. Under the PO model, Martinez et al. (2017) showed that the hypothesis of equivalence of two survival functions can be formulated as a statistical

hypothesis involving only the survival odds ratio parameter θ. Specifically, the H_0 and H_a in (1.1) are equivalent to the statement

$$\tilde{H}_0^\star : \theta \leq \frac{1}{1+\epsilon} \text{ or } \theta \geq 1+\epsilon, \text{ versus } \tilde{H}_a^\star : \frac{1}{1+\epsilon} < \theta < 1+\epsilon,$$

where $\epsilon = 4\delta/(1-\delta^2)$.

Under the model

$$\frac{1-S_1(t)}{S_1(t)} = \theta e^{z^T\gamma}\left[\frac{1-S_0(t)}{S_0(t)}\right]$$

they further demonstrated that their proposed tests are applicable even in the presence of additional covariates z beyond treatment arms. Their simulation study indicate that the proposed test procedures have correct type I error rates under the PO model as well as the PH model. Thus, their proposed tests can be a more robust practice for equivalence trials of survival responses than the commonly used log-rank based tests.

However, in some situations, both PH and PO assumptions can be violated. Furthermore, one disadvantage of PH or PO models is that it does not allow time-varying coefficients. A useful and flexible alternative is the Aalen's additive risk model (Aalen (1980, 1989, 1993), McKeague (1988); and Huffer and McKeague (1991)). This model has the feature that the influence of each covariate can vary separately and nonparametrically through time, which allows for greater flexibility of temporal structure compared with PH and PO models. In this article, we consider equivalence tests under partly Aalen's additive hazards (McKeague and Sasieni (1994)), where the influence of some covariates varies nonparametrically over time, and that of the remaining covariates containing the indicator of treatment is constant. The hazard function of partly Aalen's model $\lambda(t|X(t), Z)$ at time t, given covariates $X(t)$ and Z has the form

$$\lambda(t|X(t), Z) = X(t)^T\alpha(t) + Z^T\beta, \tag{1.2}$$

$$X(t)^T = [X_1(t), X_2(t), \ldots, X_q(t)]$$

$$Z^T = [Z_1, \ldots, Z_p]$$

are $q \times 1$ and $p \times 1$ vectors of covariates respectively,

$$\alpha(t) = [\alpha_1(t), \ldots, \alpha_q(t)]^T$$

$$\beta = [\beta_1, \ldots, \beta_p]^T$$

are the corresponding vectors of coefficients. The impact of the covariates in $X(t)$ can vary with time, but that of Z is restricted to be constant. The first component of $X(t)$, i.e., $X_1(t)$, can be set to 1 to allow for a general baseline hazard.

In Section 2, we develop an equivalence test for the difference of two hazard rates under model (1,1). In Section 3, simulation studies are conducted to investigate the finite sample performance of the proposed test.

2 The Proposed Test

Let T denote failure time and C denote the censoring time, which is assumed to be independent of T conditional on Z. Let $\tilde{X}(t) = \{X(s), s \in (0,t)\}$. The survival function $P(T > t|\tilde{X}(t), Z) = S(t|\tilde{X}(t), Z)$ is identifiable on $[0, \tau]$, where τ denotes the end point of the study. Suppose a clinical trial consists of two independent groups labeled "1" and "2", where group 1 is the standard treatment (control) group and group 2 is the new treatment group. When the treatment effect is time-invariant, we can without loss of generality denote Z_1 as an indicator variable representing treatment group membership with $Z_1 = 0$ for the standard treatment group and $Z_1 = 1$ for the new treatment group. Let $Z^* = (Z_2, \ldots, Z_p)^T$. We will now use the first element of Z, i.e., Z_1, as an indicator variable representing treatment group membership with $Z_1 = 0$ for the standard treatment group and $Z_1 = 1$ for the new treatment group. Let $S_0(t|\tilde{X}(t), Z^*)$ and $S_1(t|\tilde{X}(t), Z^*)$ denote the survival functions for standard treatment and new treatment and $\Lambda_0(t|\tilde{X}(t), Z^*)$ and $\Lambda_1(t|\tilde{X}(t), Z^*)$ are the corresponding cumulative hazard functions. In stead of considering the difference between $S_0(t|\tilde{X}(t), Z^*)$ and $S_1(t|\tilde{X}(t), Z^*)$, we consider two treatments are clinically equivalent if $|\Lambda_1(t|\tilde{X}(t), Z^*) - \Lambda_0(t|\tilde{X}(t), Z^*)|$ the difference between two cumulative hazard functions, is smaller than a predetermined equivalence level Δ over time. Thus, two treatment arms are equivalent only when

$$|\Lambda_1(t|\tilde{X}(t), Z^*) - \Lambda_0(t|\tilde{X}(t), Z^*)| < \Delta \ for \ all \ t,$$

Due to identifiability, we consider testing the null hypothesis

$$H_0 : \sup_{t\in[0,\tau]} |\Lambda_1(t|\tilde{X}(t), Z^*) - \Lambda_0(t|\tilde{X}(t), Z^*)| \geq \Delta$$

versus

$$H_a : \sup_{t\in[0,\tau]} |\Lambda_1(t|\tilde{X}(t), Z^*) - \Lambda_0(t|\tilde{X}(t), Z^*)| < \Delta$$

Under model (1.2),

$$\Lambda_1(t|\tilde{X}(t), Z^*) = \int_0^t X(s)^T \alpha(s) ds + Z^{*^T} \beta^* + \beta_1 t$$

and

$$\Lambda_0(t|\tilde{X}(t), Z^*) = \int_0^t X(s)^T \alpha(s) ds + Z^{*^T} \beta^*,$$

where $\beta^* = [\beta_2, \ldots, \beta_p]^T$.

Hence, this is equivalent to testing

$$H_0^* : \sup_{t\in[0,\tau]} |\beta_1 t| \geq \Delta$$

versus

$$H_a^* : \sup_{t\in[0,\tau]} |\beta_1 t| < \Delta$$

i.e.,

$$H_0^* : |\beta_1| \geq \eta \text{ versus } H_a^* : |\beta_1| < \eta, \tag{2.1}$$

where $\eta = \Delta/\tau$.

Denote by $(x_i(t), z_i, T_i, \delta_i)$ the observed covariates $x_i(t)$, and z_i, possibly censored failure time T_i and censoring indicator δ_i for the i^{th} observation of n individuals. McKeague and Sasieni (1994) derived estimator for β and $A(t) = \int_0^t \alpha(s)ds$. We briefly describe their approach as follows. The likelihood function is

$$L(\beta, \lambda) = \prod_{i=1}^{n} \left[\lambda_i(T_i)^{\delta_i} \times \exp\left\{ - \int_0^\tau I_{[T_i \geq t]} \lambda_i(t) dt \right\} \right],$$

The log-likelihood is

$$l(\beta, \lambda) = \sum_{i=1}^{n} \left\{ \delta_i \log \lambda_i(T_i) - \int_0^\tau I_{[T_i \geq t]} \lambda_i(t) dt \right\}, \tag{2.2}$$

where

$$\lambda_i(t) = \lambda(t|x_i(t), z_i) = x_i(t)^T \alpha(t) + z_i^T \beta.$$

Differentiate $l(\beta, \lambda)$ with respect to β to obtain the parametric score function $\dot{l}_\beta$. Setting $\dot{l}_\beta = 0$ yields

$$\beta = \left(\int_0^\tau Z(t)^T W(t) Z(t) dt \right)^{-1} \left(\int_0^\tau Z(t)^T W(t) dN(t) - \int_0^\tau Z(t)^T W(t) V(t) dA(t) \right), \tag{2.3}$$

where

$$Z(t) = (z_1 I_{[T_i \geq t]}, \ldots, z_n I_{[T_n \geq t]})^T$$

$$V(t) = (x_1(t) I_{[T_i \geq t]}, \ldots, x_n(t) I_{[T_n \geq t]})^T,$$

where $W(t) = \text{diag}\{1/\lambda_i(t)\}$ is a diagonal matrix with element $1/\lambda_i(t)$,

$$N(t) = (N_1(t), \ldots, N_n(t))^T$$

where $N_i(t) = I_{[T_i \le t, \delta_i = 1]}$ is the counting process for the failure of individual i.

Next, they considered a submodel

$$\alpha(t) = \alpha(t;\zeta) = \alpha_0(t) + \zeta b(t)$$

where ζ is a one-dimensional parameter and $b(t)$ is a given q-vector of functions. Differentiate (2.2) with respect to ζ to obtain score function

$$\dot{l}_\zeta b = \int_0^\tau b^T V(t)^T W(t) dN(t) - \int_0^\tau b^T V(t)^T W(t) Z(t) \beta dt - \int_0^\tau b^T V(t)^T W(t) V(t) dA(t).$$

Setting $\dot{l}_\zeta b = 0$ for all vector valued function b, we obtain

$$A(t) = \int_0^t (V(s)^T W(s) V(s))^{-1} V(s)^T W(s) dN(s) - \int_0^t (V(s)^T W(s) V(s))^{-1} V(s)^T W(s) Z(s) \beta ds. \tag{2.4}$$

Substituting the right-hand side of (2.4) into (2.3) and solving for β gives

$$\beta_w = \left(\int_0^\tau Z(t)^T H(t) Z(t) dt \right)^{-1} \int_0^\tau Z(t)^T H(t) dN(t),$$

where

$$H(t) = W(t) - W(t)V(t)(V(t)^T W(t) V(t))^{-1} V(t)^T W(t).$$

Notice that β_w is not an estimator since $W(t)$ depends on the unknown hazard density function $\lambda_i(t)$. An ordinary least squares (OLS) estimator $\hat{\beta}$ can be obtained by replacing $W(t)$ by I, i.e,

$$\hat{\beta} = \left(\int_0^\tau Z(t)^T (I - P(t)) Z(t) dt \right)^{-1} \int_0^\tau Z(t)^T (I - P(t)) dN(t),$$

where

$$P(t) = V(t)(V(t)^T V(t))^{-1} V(t)^T$$

Based on $\hat{\beta}$, we can obtain an OLS estimator of $A(t)$, given by

$$\hat{A}(t) = \int_0^t (V(s)^T V(s))^{-1} V(s)^T dN(s) - \int_0^t (V(s)^T V(s))^{-1} V(s)^T Z(s) \hat{\beta} ds.$$

The estimator $\hat{\beta}$ is consistent and $\sqrt{n}(\hat{\beta} - \beta)$ converges in distribution to a p-variate normal with mean zero and covariance matrix Σ^{-1}, where

$$\Sigma = n^{-1} \int_0^\tau Z(t)^T (I - P(t)) Z(t) dt$$

Similarly, $n^{1/2}(\hat{A}(t) - A(t))$ converges in distribution to a q-variate Gaussian process with mean zero and covariance function which, as a function of s and t, can be consistently estimated by $n \sum_{r \leq s \wedge t} J_r J_r^T \hat{\psi}(s) \Sigma^{-1} \hat{\psi}(t)^T$,
where J_r is the jump in $\hat{A}$ at time r and

$$\hat{\psi}(s) = \int_0^s (V(u)^T V(u))^{-1} V(u)^T Z(u) ds.$$

Notice that the estimators $\hat{\beta}$ and $\hat{A}$ are not efficient estimators. McKeague and Sasieni (1994) proposed efficient estimators for β and A based on the following method:
(i) Obtain an estimator $\hat{W}(t)$ from a predictable kernel smoother, following Huffer & McKeague (1991).
(ii) Find an estimator $\hat{\beta}_w$ for β using $\hat{W}(t)$:

$$\hat{\beta}_w = \left(\int_0^\tau Z(t)^T \hat{H}(t) Z(t) dt \right)^{-1} \int_0^\tau Z(t)^T \hat{H}(t) dN(t),$$

where

$$\hat{H}(t) = \hat{W}(t) - \hat{W}(t) V(t) (V(t)^T \hat{W}(t) V(t))^{-1} V(t)^T \hat{W}(t).$$

(iii) Find an estimate $\hat{A}_w(t)$ of $A(t)$ using $\hat{W}(t)$ and $\hat{\beta}_w$:

$$\hat{A}_w(t) = \int_0^t (V(s)^T \hat{W}(s) V(s))^{-1} V(s)^T \hat{W}(s) dN(s) - \int_0^t (V(s)^T \hat{W}(s) V(s))^{-1} V(s)^T \hat{W}(s) Z(s) \hat{\beta}_w ds.$$

The estimator $\hat{\beta}_w$ is consistent and $\sqrt{n}(\hat{\beta}_w - \beta)$ converges in distribution to a p-variate normal with mean zero and covariance matrix, which can be consistently estimated by

$$\hat{\Sigma}_w = n^{-1} \int_0^\tau Z(t)^T \hat{H}(t) Z(t) dt.$$

Based on $\hat{\beta}$ and $\hat{\beta}_w$, we propose two Wald type tests to test the null hypothesis

$$H_0^* : |\beta_1| \geq \eta.$$

Let $se(\hat{\beta})$ denote the asymptotic standard error of $\hat{\beta}$, which can be obtained from Σ. Similarly, let $se(\hat{\beta}_w)$ denote the estimated asymptotic standard error of $\hat{\beta}_w$, which can be calculated from $\hat{\Sigma}_w$. We consider the following two testing statistics:

$$\hat{T} = \frac{|\hat{\beta}_1|}{se(\hat{\beta}_1)} \text{ and } \hat{T}_w = \frac{|\hat{\beta}_{1w}|}{se(\hat{\beta}_{1w})},$$

where $\hat{\beta}_1$ and $\hat{\beta}_{1w}$ are the first element of $\hat{\beta}$ and $\hat{\beta}_w$, respectively, and $se(\hat{\beta}_1)$ and $se(\hat{\beta}_{1w})$ are their corresponding standard errors. The rejection regions of the two tests are

$$\hat{T} < C_\alpha(\eta/se(\hat{\beta}_1)$$

and

$$\hat{T}_w < C_\alpha(\eta/se(\hat{\beta}_{1w})$$

respectively, where the $C_\alpha^2(\psi)$ is the α^{th} quantile of a χ^2 distribution with degree of freedom equal to 1 and non centrality parameter ψ.

Since $se(\hat{\beta}_1)$ is based on the asymptotic results, it can severely underestimate the standard deviation. One alternative is to consider the Jackknife method. The jackknife technique is well described in Mosteller and Tukey (1977) and has been shown to be widely useful for obtaining robust confidence intervals. For randomly censored data, Gaver and Miller (1983) demonstrated that the Kaplan-Meier survival estimator can be jackknifed to give conservative confidence limits for survival probability. Here, we consider delete-one jackknife estimate of standard error, denoted by $se_J(\hat{\beta}_1)$ and construct testing statistics based on

$$\hat{T}_J = \frac{|\hat{\beta}_1|}{se_J(\hat{\beta}_1)}.$$

3 Simulation Study

A simulation study is conducted to investigate the performance of the proposed tests. We consider the simulation model

$$\lambda(t|x(t), z) = 1 + x_{2i}(t)^T \alpha_2(t) + \beta_1 z_1,$$

where intercept is $\alpha_1(t) = 1$, $\alpha_2(t) = t$, and $x_{2i}(t)$'s are generated from discrete distribution with $P(x_{2i}(t) = j) = 0.25$ for $j = 1, 2, 3, 4$ and $z_1(t)$ is a Bernoulli random variable with success probability equal to 0.5. The value of η is set as 0.8 and the values of β_1 are set as 0.2, 0.4, 0.6 and 0.8. The right censoring variable is generated as $C = \min(C_1, C_2)$, where C_1 is a constant and equal to one and C_2 was generated from exponential distribution with means 1.25 and 2.5 such that censoring rates are about 0.30 and 0.20, respectively. Sample size is set at 100, 200 and 400 and the replication times is 1000. Table 1 and Table 2 show the biases of the estimates $\hat{\beta}$, and $\hat{\beta}_w$, their asymptotic estimated standard error (denoted by se), and delete-one jackknife estimate of standard error (denoted by se_J) for $\hat{\beta}$. Tables 1 and 2 also show the testing powers based on testing statistics $\hat{T}$, $\hat{T}_w$ and $\hat{T}_J$ with Type I error at $\alpha = 0.05$. Table 3 and Table 4 show the biases and standard deviations of $\hat{A}_1(t)$, $\hat{A}_2(t)$, $\hat{A}_{1w}(t)$, and $\hat{A}_{2w}(t)$ at $t = 0.7, 0.8$ and 0.9.

Based on Tables 1 through 4, we have the following conclusions:

(i) Table 1 and 2 indicate that the estimated asymptotic standard errors of $\hat{\beta}$ severly underestimate the true standard errors while the delete-one jackkinfe estimates are close to the true values. Thus, the type I error rates of the testing statistics $\hat{T}$ are larger than the nominal level $\alpha = 0.05$ while that of $\hat{T}_j$ are close to nominal level.

(ii) Since the estimated asymptotic standard errors of $\hat{\beta}_w$ are close to the true standard errors, the testing statistics $\hat{T}_w$ performs well.

(iii) Given sample size n, the power of all the three tests increase as the values of β_1 decrease. Give β_1, the power of all the three tests increases as sample size increases.

(iv) Tables 3 and 4 indicate that both $\hat{A}_i(t)$ and $\hat{A}_{wi}(t)$ perform well with small biases. When $n = 400$, $\hat{A}_{wi}(t)$ is more efficient than $\hat{A}_i(t)$ but the efficiency gain is limited.

Table 1. Simulation results for $\hat{\beta}_1$, $\hat{T}$, $\hat{T}_J$, $\hat{\beta}_{1w}$, $\hat{T}_w$, $\eta = 0.8$

C_2: exponential distribution with means 1.25

			$\hat{\beta}_1$			$\hat{T}$	$\hat{T}_J$	$\hat{\beta}_{1w}$			$\hat{T}_w$
β	n	bias	ese	se_J	se	power	power	bias	ese	se	power
0.2	100	0.012	0.339	0.506	0.438	0.380	0.197	0.009	0.430	0.448	0.232
0.4	100	0.017	0.344	0.530	0.476	0.266	0.129	0.039	0.451	0.469	0.151
0.6	100	0.036	0.350	0.551	0.499	0.175	0.067	0.016	0.472	0.487	0.099
0.8	100	0.001	0.357	0.571	0.510	0.123	0.044	0.029	0.494	0.495	0.059
1.0	100	0.037	0.363	0.595	0.548	0.057	0.031	0.015	0.516	0.521	0.022
0.2	200	0.006	0.236	0.343	0.314	0.715	0.458	-0.004	0.299	0.294	0.613
0.4	200	0.017	0.240	0.356	0.333	0.476	0.258	-0.004	0.314	0.319	0.356
0.6	200	0.013	0.244	0.373	0.346	0.261	0.103	0.018	0.329	0.346	0.155
0.8	200	0.016	0.248	0.389	0.358	0.127	0.047	0.025	0.344	0.352	0.041
1.0	200	0.007	0.252	0.403	0.378	0.045	0.008	0.008	0.359	0.360	0.017
0.2	400	0.005	0.165	0.238	0.234	0.933	0.825	0.015	0.210	0.211	0.869
0.4	400	0.002	0.168	0.247	0.229	0.704	0.480	0.010	0.220	0.226	0.541
0.6	400	0.007	0.171	0.256	0.238	0.376	0.205	0.004	0.230	0.231	0.217
0.8	400	0.005	0.173	0.268	0.249	0.120	0.054	0.006	0.241	0.244	0.042
1.0	400	0.001	0.176	0.279	0.264	0.030	0.010	-0.001	0.252	0.253	0.007

◎ When $\beta = 0.8$, power is the estimated Type I error

Table 2. Simulation results for $\hat{\beta}_1$, $\hat{T}$, $\hat{T}_J$, $\hat{\beta}_{1w}$, $\hat{T}_w$, $\eta = 0.8$

C_2: exponential distribution with means 2.5

			$\hat{\beta}_1$			$\hat{T}$	$\hat{T}_J$	$\hat{\beta}_{1w}$			$\hat{T}_w$
β	n	bias	ese	se_J	se	power	power	bias	ese	se	power
0.2	100	-0.022	0.318	0.465	0.450	0.447	0.197	0.004	0.408	0.411	0.295
0.4	100	0.001	0.324	0.486	0.467	0.332	0.163	0.019	0.428	0.450	0.173
0.6	100	0.014	0.329	0.505	0.482	0.189	0.079	0.012	0.448	0.458	0.094
0.8	100	0.026	0.336	0.530	0.506	0.117	0.049	0.022	0.469	0.483	0.047
1.0	100	0.032	0.342	0.552	0.539	0.051	0.019	0.035	0.492	0.521	0.025
0.2	200	-0.013	0.221	0.316	0.298	0.788	0.549	0.006	0.285	0.278	0.664
0.4	200	-0.007	0.226	0.330	0.306	0.537	0.318	-0.011	0.299	0.309	0.406
0.6	200	-0.001	0.229	0.344	0.329	0.314	0.135	0.013	0.313	0.313	0.159
0.8	200	0.023	0.233	0.356	0.348	0.134	0.043	0.010	0.327	0.328	0.043
1.0	200	0.016	0.238	0.374	0.355	0.039	0.009	0.019	0.343	0.353	0.012
0.2	400	0.002	0.156	0.219	0.203	0.938	0.860	0.001	0.200	0.199	0.909
0.4	400	0.002	0.158	0.227	0.216	0.692	0.585	0.006	0.210	0.208	0.588
0.6	400	-0.001	0.161	0.238	0.226	0.405	0.210	0.007	0.220	0.222	0.219
0.8	400	-0.003	0.164	0.248	0.236	0.111	0.054	0.006	0.230	0.236	0.049
1.0	400	-0.005	0.167	0.259	0.261	0.034	0.006	0.001	0.241	0.251	0.006

◎ When $\beta = 0.8$, power is the estimated of Type I error

Table 3. Simulation results for $\hat{A}_1(t)$, $\hat{A}_2(t)$, $\hat{A}_{1w}(t)$, $\hat{A}_{2w}(t)$
C_2: exponential distribution with means 1.25

				$\hat{A}_1(t)$		$\hat{A}_2(t)$		$\hat{A}_{1w}(t)$		$\hat{A}_{2w}(t)$	
t	n	$A_1(t)$	$A_2(t)$	bias	se	bias	se	bias	se	bias	se
0.7	100	0.7	0.245	0.015	0.552	-0.002	0.226	0.008	0.527	0.001	0.217
0.7	200	0.7	0.245	0.006	0.381	-0.001	0.154	-0.014	0.366	0.008	0.154
0.7	400	0.7	0.245	0.002	0.264	0.001	0.113	0.003	0.243	-0.006	0.102
0.8	100	0.8	0.320	-0.001	0.717	0.008	0.310	-0.020	0.682	-0.001	0.293
0.8	200	0.8	0.320	0.032	0.477	-0.008	0.205	0.019	0.443	-0.004	0.189
0.8	400	0.8	0.320	-0.012	0.316	0.002	0.139	-0.016	0.290	-0.009	0.131
0.9	100	0.9	0.405	-0.047	0.858	0.030	0.415	-0.077	0.854	0.046	0.414
0.9	200	0.9	0.405	0.027	0.562	-0.009	0.245	0.014	0.567	0.004	0.256
0.9	400	0.9	0.405	0.003	0.379	0.004	0.175	0.010	0.373	0.001	0.171

Table 4. Simulation results for $\hat{A}_1(t)$, $\hat{A}_2(t)$, $\hat{A}_{1w}(t)$, $\hat{A}_{2w}(t)$
C_2: exponential distribution with means 2.5

				$\hat{A}_1(t)$		$\hat{A}_2(t)$		$\hat{A}_{1w}(t)$		$\hat{A}_{2w}(t)$	
t	n	$A_1(t)$	$A_2(t)$	bias	se	bias	se	bias	se	bias	se
0.7	100	0.7	0.245	0.030	0.503	-0.010	0.205	0.022	0.502	-0.003	0.204
0.7	200	0.7	0.245	0.010	0.332	-0.001	0.136	0.009	0.325	-0.003	0.133
0.7	400	0.7	0.245	0.006	0.227	-0.004	0.095	0.012	0.239	-0.003	0.092
0.8	100	0.8	0.320	-0.010	0.629	0.015	0.273	-0.006	0.573	0.009	0.259
0.8	200	0.8	0.320	0.002	0.402	0.007	0.176	0.005	0.390	-0.005	0.165
0.8	400	0.8	0.320	0.016	0.288	-0.003	0.121	0.014	0.273	-0.007	0.115
0.9	100	0.9	0.405	-0.010	0.815	0.015	0.362	0.007	0.744	0.005	0.353
0.9	200	0.9	0.405	0.017	0.475	-0.003	0.217	0.018	0.483	-0.009	0.220
0.9	400	0.9	0.405	0.009	0.331	-0.002	0.152	0.001	0.308	0.003	0.139

4 Discussion

When the treatment effect is time-dependent, we can without loss of generality denote $X_2(t) \equiv X_2$ as an indicator variable representing treatment group membership with $X_2 = 0$ for the standard treatment group and $X_2 = 1$ for the new treatment group. Let $X^*(t) = (X_3(t), \ldots, X_q(t))^T$ and $\tilde{X}^*(t) = \{s : X^*(s), s \in (0, t)\}$. Let $\Lambda_0(t|\tilde{X}^*(t), Z)$ and $\Lambda_1(t|\tilde{X}^*(t), Z)$ denote the cumulative hazard functions for standard treatment and new treatment. Thus, two treatment arms are equivalent only when

$$|\Lambda_1(t|\tilde{X}^*(t), Z) - \Lambda_0(t|\tilde{X}^*(t), Z)| < \Delta \ for\ all\ t.$$

Due to identifiability, we consider testing the null hypothesis

$$H_0 : \sup_{t\in[0,\tau]} |\Lambda_1(t|\tilde{X}^*(t), Z) - \Lambda_0(t|\tilde{X}^*(t), Z)| \geq \Delta$$

versus

$$H_a : \sup_{t\in[0,\tau]} |\Lambda_1(t|\tilde{X}^*(t), Z) - \Lambda_0(t|\tilde{X}^*(t), Z)| < \Delta.$$

Under model (1.1), this is equivalent to testing

$$H_0^* : \sup_{t\in[0,\tau]} |A_2(t)| \geq \Delta$$

versus

$$H_a^* : \sup_{t\in[0,\tau]} |A_2(t)| < \Delta.$$

Further resarch is required to deveolp testing statistics to deal with this case.

References

Aalen, O. O. (1980). A model for non-parametric regression analysis of counting process. In:Klonecki, W., Kozek, A., Rosiski, J. (Eds). Lecture Notes on Mathematical Statistics and Probability, Vol. 2. Springer, New York, pp. 1-25.

Aalen, O. O. (1989). A linear regression model for analysis of life times. *Statistics in Medicine*, **8**, 907-925.

Aalen, O. O. (1993). Further results on the non-parametric linear regression model in survival analysis. *Statistics in Medicine*, **17**, 1569-1588.

Bennett, S. (1983). Analysis of survival data by the proportional odds model. *Statistics in Medicine*, **2**, 273-277.

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

Huffer, F. W. and Mckeague, F. W. (1991). Weighted least squares estimation for Aalen's additive risk model. *Journal of the American Statistical Association*, **86**, 114-129.

Martines, E. E., Sinha, D. Wang, W. Lipsitz, S. R., Chappell, R. J. (2017). Tests for equivalence of two survival functions: Alternative to the tests under proportional hazards. *Statistical Methods in Medical Research*, **26**, 75-87.

McKeagues, I. W. (1988). Asymptotic theory for weighted least squares estimators in Aalen's additive risk model. *in Statistical Inference from Stochastic Processes*, page 139-152.

McKeagues, I. W. and Sasieni, P. D. (1994). A partly parametric additive risk model. *Biometrika*, **81**, 501-514.

Murphy, S. A., Rossini, A. J. and van der Vaart, A. W. (1997). Maximum likelihood estimation in the proportional odds model. *Journal of the American Statistical Association*, **92**, 968-976.

Sasieni, P. D. (1992). Information bounds for the additive and multiplicative intensity models. In: *Survival Analysis of the art*, (J.P. Klein and P.K. Goel, eds.),

249-265, Kluwer.

Tukey, J. (1958). Bias and confidence in not quite large samples, *The Annals of the institute Mathematical Statistic*, **29**, 614.

Wellek, S. (1993). A log-rank test for the equivalence of two survival functions. *Biometrics*, **49**, 877-881.