Published in final edited form as:

Lifetime Data Anal. 2004 June ; 10(2): 103–120.

Two-Sample Statistics for Testing the Equality of Survival Functions Against Improper Semi-parametric Accelerated Failure Time Alternatives: An Application to the Analysis of a Breast Cancer Clinical Trial

PHILIPPE BROËT,

Department of Public Health and National Institute for Health and Medical Research (INSERM U472) Paul Brousse Hospital, Villejuif, France and Institut Curie, 26 rue d'Ulm, Paris, France
broet@vjf.inserm.fr

ALEXANDER TSODIKOV,

Huntsman Cancer Institute 2000 Circle of Hope, Salt Lake City, Utah 84112-5550, USA

YANN DE RYCKE, and

Institut Curie, 26 rue d'Ulm, Paris, France

THIERRY MOREAU

National Institute for Health and Medical Research (INSERM U472) Paul Brousse Hospital, Villejuif, France

Abstract

This paper presents two-sample statistics suited for testing equality of survival functions against improper semi-parametric accelerated failure time alternatives. These tests are designed for comparing either the short- or the long-term effect of a prognostic factor, or both. These statistics are obtained as partial likelihood score statistics from a time-dependent Cox model. As a consequence, the proposed tests can be very easily implemented using widely available software. A breast cancer clinical trial is presented as an example to demonstrate the utility of the proposed tests.

Keywords

accelerated failure time models; cure rate model; improper model; semi-parametric model

1. Introduction

In recent years, there has been a renewed interest in methods for analyzing survival data with long-term survivors fraction or a 'cure fraction' (for a review, see Maller and Zhou, 1996). Most of these methods attempt to distinguish between the different mechanisms by which a prognostic factor may act on the event's occurrence. Indeed, a prognostic factor may affect either the probability of never experiencing the event of interest (termed 'long-term effect' in the following text) or the time to occurrence of the event (termed 'short-term effect' in the following text), or both.

Testing procedures for these effects were proposed in a recent paper where we assumed proportional hazards for the short-term effect (Broët et al., 2001). In the present paper, we extend this procedure to a non-proportional hazards behavior of the short-term effect. This work was motivated by the analysis of a breast cancer randomized trial comparing the

distribution of the disease-free interval in two treatments groups where Kaplan–Meier curves (Kaplan and Meier, 1958) cross each other during the follow-up (see Figure 1). The two groups are defined according to the different way of administration of the same chemotherapy which is scheduled either to follow (adjuvant) or precede (primary or neo-adjuvant) the local regional treatment. The question addressed in Section 5 is whether primary chemotherapy, which shrinks the tumor before local treatment, modifies the timing of the recurrence as compared to adjuvant chemotherapy, taking a long-term recurrence-free rate into account. This situation was quite puzzling and prompted us to derive test statistics suited for this case. As it will be seen, the proposed tests provide interesting results that would have been overlooked if only results from classical tests were considered.

In the literature, the most common approach for modeling failure-time data with a cure fraction relies on the assumption that the overall distribution of the survival times is a mixture of two components: one corresponding to the subjects who are not susceptible ('cured' subjects) and the other to the subjects who are susceptible of experiencing the event ('uncured' subjects). In this setting, most of the published non-parametric procedures which are convenient for the two-sample comparison do not allow testing for both effects (short-term and long-term) (Gray and Tsiatis, 1989; Laska and Meisner, 1992; Sposto et al. 1992; Lee, 1995). Others require complex computations which are too heavy for a practical use in routine (Kuk and Chen, 1992; Taylor, 1995).

A different approach which overcomes these drawbacks relies on models that define the cumulative hazard as a bounded increasing positive function in a parametric (Aalen, 1992; Cantor and Shuster, 1992; Yakovlev and Tsodikov, 1996) or semi-parametric way (Tsodikov, 1998; Shen and Sinha, 2002; Ibrahim, 1999, 2001; Broët et al., 2001). Such semi-parametric modeling was used in our previous work to test for no short-term or no long-term effect against improper short-term proportional hazard alternatives (Broët et al., 2001). In this paper, the restrictive proportional hazard short-term effect assumption is relaxed and statistics are proposed for improper short-term accelerated failure time alternatives.

In Section 2, a semi-parametric improper accelerated failure time model is described. In Section 3, the proposed score statistics are derived from a Cox model with a time-dependent covariate. In Section 4, we present the results of simulation experiments. In Section 5, the clinical relevance of these tests is demonstrated by the analysis of a breast cancer clinical trial with long-term follow-up. Section 6 contains a discussion and guidelines for the use of the tests.

2. The Semi-Parametric Improper Accelerated Failure Time Model

Let $i = 0, 1$ denote the two groups to be compared, with n_i subjects in group i ($n = n_0 + n_1$). For each patient j , let the random variables T_j and C_j be the survival and censoring times which are assumed to satisfy the condition of independent censoring (Fleming and Harrington, 1991, pp. 26–27). We denote $X_j = \min(T_j, C_j)$ the observed time of follow-up, $\delta_j = 1_{\{X_j = T_j\}}$ the indicator of death, $Y_j(t) = 1_{\{t \leq X_j\}}$ the indicator of being at risk at time t , and Z_j the indicator variable of group 1. For the subject j , the data consist of X_j , δ_j and Z_j . The hazard function of T_j corresponding to every subject j belonging to group i is denoted by: $\lambda_i(t) = f_i(t)/S_i(t)$, where $f_i(t)$ and $S_i(t)$ are the probability density function and the survival function, respectively. The corresponding cumulative hazard function is denoted by $\Lambda_i(t) = -\log[S_i(t)]$.

A semi-parametric improper model is defined by the following general survival function in group i :

$$S_i(t) = \exp \left\{ -\theta e^{\beta_1 i} [1 - A(t, \beta_2 i)] \right\} \quad (1)$$

where $A(t, \beta_2 i)$ is a function decreasing with time from one to zero, which is similar to a survival function, and where θ is a positive parameter. The function $S_i(t)$ is improper and its limiting value $\exp(-\theta e^{\beta_1 i})$ is called the tail defect and represents the probability of not experiencing the event of interest in group i . The cumulative hazard $\Lambda_i(t) = \theta e^{\beta_1 i} [1 - A(t, \beta_2 i)]$ is less than or equal to $\theta e^{\beta_1 i}$.

The model (1) has two components: the first term containing β_1 which quantifies the long-term effect and the function $A(t, \beta_2 i)$ which expresses the short-term effect. More precisely, if $\beta_1 = 0$, the two groups have the same cure fraction (no long-term effect) and if $\beta_2 = 0$, the model reduces to a proportional hazard model. In this case, the relative risk is constant over time which implies no short-term effect.

A particular case of (1) was considered in a previous work (Broët et al., 2001), where we assumed a proportional hazard modeling of the short-term effect, so that in the general formulation (1) given here, $A(t, \beta_2 i) = A(t) e^{\beta_2 i}$. Here, we consider another particular case of (1) where a non-proportional hazards model is assumed for the short-term effect by letting: $A(t, \beta_2 i) = \exp[-K(t) e^{\beta_2 i}]$ where $K(t)$ is a positive function increasing with time from zero to infinity. This is an obvious semi-parametric generalization of the case of two Weibull distributions differing in their shape parameters. The resulting model

$$S_i(t) = \exp \left\{ -\theta e^{\beta_1 i} \left[1 - \exp(-K(t) e^{\beta_2 i}) \right] \right\} \quad (2)$$

has the following property. In case of no long-term effect ($\beta_1 = 0$) and with a short-term effect such as $\beta_2 < 0$, the survival functions $S_0(t)$ and $S_1(t)$ cross before converging to the same long-term survivor fraction.

3. Proposed Test Statistics

In this section, statistics are derived for testing ($\beta_1 = 0$) and/or ($\beta_2 = 0$) in model (2). The derivation is achieved by using a proportional hazards model with a time-dependent covariate which approximates (2) about ($\beta_2 = 0$) and which serves as a basis for computing the desired statistic. They are easily computed as score statistics from the usual partial likelihood. The null hypotheses to be tested are: $H_0: (\beta_1 = \beta_2 = 0)$; $H_{00}: (\beta_2 = 0)$ and $H_{000}: (\beta_1 = 0)$.

3.1. Method for Deriving the Test Statistics

Now, we define the following quantity: $D(t, \beta_2 i) = -\frac{\partial}{\partial t} A(t, \beta_2 i)$ which refers to the density function related to $A(t, \beta_2 i)$. The general model (1) can be written in terms of the hazard functions $\lambda_0(t)$ and $\lambda_1(t)$:

$$\log [\lambda_1(t) / \lambda_0(t)] = \beta_1 + \log [D(t, \beta_2) / D(t, 0)] \quad (3)$$

Expanding $\log(D(t, \beta_2))$ about $\beta_2 = 0$ in (3) gives the following first-order approximation:

$$\log [\lambda_1(t) / \lambda_0(t)] = \beta_1 + \beta_2 w(t) \quad (4)$$

with

$$w(t) = \left[\frac{\partial}{\partial \beta_2} \log D(t, \beta_2) \Big|_{\beta_2=0} \right]$$

Under the improper short-term accelerated failure time model (2), $w(t)$ is equal to

$$w(t) = 1 + \log[-\log A(t)] + [\log A(t)] \log[-\log A(t)]$$

where

$$A(t) = \left[1 - \frac{\Lambda_0(t)}{\theta} \right]. \quad (5)$$

In case of improper short-term proportional hazard model, $w(t) = 1 + [\log A(t)]$. Substituting $\Lambda_0(t)$ and θ in (5) by efficient estimators under the null hypothesis to be tested provides estimates $\hat{w}(t)$ of $w(t)$. These estimators are presented for each null hypothesis H_0 , H_{00} and H_{000} in the next subsection. Replacing $w(t)$ by $\hat{w}(t)$ in (4), defines a time-dependent proportional hazards model with the internal time-dependent covariate $\hat{w}(t)$ (Kabfleich and Prentice, 1980). The proposed statistics for testing $(\beta_1 = 0)$ and/or $(\beta_2 = 0)$ can be easily derived as the score statistics from this time-dependent proportional hazards model through the corresponding partial likelihood. It can be easily shown that the resulting score statistics for testing the lack of short-term effect with or without a long-term effect are the same as in model (2), which would not be the case for the likelihood ratio or Wald tests. Moreover, the proposed score statistic for testing for no long-term effect can be easily derived while similar derivation from model (2) would be at least burdensome.

The resulting score statistics depend on the unknown parameters $\Lambda_0(t)$ and θ . Replacing $\Lambda_0(t)$ and θ by efficient estimators and applying the results of Pierce (Pierce, 1982) to our setting as presented in an earlier work (Broët et al., 2001) for improper short-term proportional hazards model allows us to obtain the asymptotic distributions of the proposed statistics.

Score statistics for testing H_0 and H_{00} are asymptotically distributed as chi-squares with two degrees and one degree of freedom, respectively. Concerning H_{000} , it should be noted that the corresponding score statistic depends on an estimate of β_2 as seen in the next section. As the score statistic is derived from model 3 (based on a first-order approximation) which is valid under $\beta_2 = 0$, the score statistic is approximately distributed as a χ^2 with one degree of freedom for small values of β_2 . This is not the case for the two other tests that do not depend on β_2 under their corresponding null hypothesis. For the validity of the results it is required that the upper bound of the domain for which the survival distribution of the survival time variable is greater than zero, be less than the upper bound of the censoring distribution. In practice, this condition expresses the fact that the susceptible subjects should experience the event within the maximum length of follow-up. It should be stressed that the distribution of the score statistics for testing H_0 and H_{00} is a chi-square distribution no matter whether the sufficient follow-up condition holds true or not. Indeed, the null hypotheses H_0 and H_{00} do not involve $A(t)$ and are identical under the two models.

3.2. Score Tests

3.2.1. Testing the Lack of Short and Long-Term effect—The components of the score vector for testing $H_0: \beta_1 = \beta_2 = 0$ can be written as follows:

$$\begin{aligned}\widehat{V}_{H_{0,1}} &= \frac{\partial \log L}{\partial \beta_1} = \sum_{j=1}^n \delta_j \left\{ Z_j - \frac{\sum_{k=1}^n Y_k(t_j) \cdot Z_k}{\sum_{k=1}^n Y_k(t_j)} \right\} \\ \widehat{V}_{H_{0,2}} &= \frac{\partial \log L}{\partial \beta_2} = \sum_{j=1}^n \delta_j \widehat{w}(t_j) \left\{ Z_j - \frac{\sum_{k=1}^n Y_k(t_j) \cdot Z_k}{\sum_{k=1}^n Y_k(t_j)} \right\}\end{aligned}$$

In $\widehat{V}_{H_{0,2}}$, $\widehat{w}(t)$ is computed as indicated in Section 3.1 using the left-continuous version of the Nelson–Aalen estimator (Nelson, 1972; Aalen, 1978) for $\widehat{\Lambda}_0(t)$ and using its value computed at the last observed failure time for $\widehat{\theta}$. The corresponding observed information matrix $\widehat{I}_{H_0}$ under H_0 is given in the Appendix.

Under H_0 , the statistic $S_{H_0} = [\widehat{V}_{H_{0,1}}, \widehat{V}_{H_{0,2}}] \widehat{I}_{H_0}^{-1} [\widehat{V}_{H_{0,1}}, \widehat{V}_{H_{0,2}}]'$ is asymptotically distributed as a chi-square with two degrees of freedom.

3.2.2. Testing the Lack of Short-Term Effect—The components of the score vector for testing $H_{00}: \beta_2 = 0$, for any β_1 can be written as follows:

$$\begin{aligned}\widehat{V}_{H_{00,1}} &= \frac{\partial \log L}{\partial \beta_1} = 0 \\ \widehat{V}_{H_{00,2}} &= \frac{\partial \log L}{\partial \beta_2} = \sum_{j=1}^n \delta_j \widehat{w}(t_j) \left\{ Z_j - \frac{\sum_{k=1}^n Y_k(t_j) e^{\widehat{\beta}_1 Z_k} Z_k}{\sum_{k=1}^n Y_k(t_j) e^{\widehat{\beta}_1 Z_k}} \right\}\end{aligned}$$

In $\widehat{V}_{H_{00,2}}$, $\widehat{V}_{H_{00,2}}, \widehat{\beta}_1$ is the usual partial likelihood estimator of β_1 under H_{00} ; $\widehat{w}(t)$ is computed by using the left-continuous version of the Breslow's estimator [Breslow, 1972, 1974] for $\widehat{\Lambda}_0(t)$ under H_{00} and for $\widehat{\theta}$ its value computed at the last observed failure time. The Breslow's estimator for $\Lambda_0(t)$ under H_{00} is given by

$$\sum_{k=1}^n \delta_k \left[\sum_{j=1}^n Y_j(t_k) \right]^{-1}.$$

The corresponding observed information matrix $\widehat{I}_{H_{00}}$ under H_{00} is given in the Appendix.

Under H_{00} , the statistic $S_{H_{00}} = [0, \widehat{V}_{H_{00,2}}] \widehat{I}_{H_{00}}^{-1} [0, \widehat{V}_{H_{00,2}}]'$ is asymptotically distributed as a χ^2 with one degree of freedom.

3.2.3. Testing the Lack of Long-Term Effect—The components of the score vector for testing $H_{000}: \beta_1 = 0$, for any β_2 can be written as follows:

$$\begin{aligned}\widehat{V}_{H_{000},1} &= \frac{\partial \log L}{\partial \beta_1} = \sum_{j=1}^n \delta_j \left\{ Z_j - \frac{\sum_{k=1}^n Y_k(t_j) e^{\widehat{\beta}_2 \widehat{w}(t_j)} Z_k}{\sum_{k=1}^n Y_k(t_j) e^{\widehat{\beta}_2 \widehat{w}(t_j)} Z_k} \right\} \\ \widehat{V}_{H_{000},2} &= \frac{\partial \log L}{\partial \beta_2} = 0\end{aligned}$$

where $\widehat{V}_{H_{000},1}$ is obtained by using $\widehat{w}(t)$ as given in Section 3.2.

A derivation of $\widehat{w}(t)$ and $\widehat{\beta}_2$ could be achieved through an iterative procedure. For computational simplicity, the first-step estimators are used in the proposed score statistic. The procedure is as follows. At first step, $\widehat{w}(t)$ is taken under H_0 where the cumulative baseline hazard is replaced by the Nelson–Aalen estimator and $\widehat{\beta}_2$ is taken as the partial likelihood estimator obtained in the corresponding time-dependent model. At second step, $\widehat{\beta}_2$ thus obtained, is used to update $\widehat{w}(t)$ using the left-continuous version of a Breslow's type estimator given by

$$\sum_{k=1}^n \delta_k \left[\sum_{j=1}^n Y_j(t_k) e^{[Z_j \widehat{\beta}_2 \widehat{w}(t_k)]} \right]^{-1}$$

for $\widehat{\Lambda}_0(t)$. As mentioned above, the proposed statistic is computed by using the first-step estimator $\widehat{\beta}_2$ together with $\widehat{w}(t)$. This implies that $\widehat{V}_{H_{000},2}$ does not vanish but is taken to be null in the computation. The corresponding observed information matrix $\widehat{I}_{H_{000}}$ is given in the Appendix.

For small values of β_2 , under H_{000} , the statistic $S_{H_{000}} = [\widehat{V}_{H_{000},1}, 0] \widehat{I}_{H_{000}}^{-1} [\widehat{V}_{H_{000},1}, 0]'$ is approximately distributed as a χ^2 with one degree of freedom.

4. Simulation Study

4.1. Method

A simulation study was performed to investigate the power properties of the proposed tests in comparison with classical tests such as the Logrank test (LR) (Peto and Peto, 1972) and the Peto–Prentice–Wilcoxon test (PPW) (Kabfleich and Prentice, 1980). The proposed tests of H_0 , H_{00} and H_{000} are denoted SLT, ST and LT, respectively. We also consider the test for no short- and no long-term effect (SLT-PH) designed for improper short-term proportional hazard alternatives [Broët et al., 2001]. In addition, the product-limit test (PL) which is a non-parametric test of no difference in the cure fraction (Spoto, Sather and Baker, 1992) is also considered.

Data were generated to mimic a simple randomized clinical trial with two different models: (A) improper short-term accelerated failure time model, (B) improper short-term proportional hazard model. Survival times were generated according to model (1) with $A(t, \beta_2 i) = \exp(-te^{\beta_2 i})$ and $A(t, \beta_2 i) = \exp(-te^{\beta_2 i})$ for the proportional hazards and accelerated failure time model, respectively. Censoring times were independently generated from a uniform distribution over $[0, u]$. It worth noting that in the uniform censoring case, a susceptible subject may not experience the event of interest within the follow-up time u . For each set of parameter values u can be easily computed so as to ensure a given percentage of censoring. The percentage of censoring refers only to the percentage of censored

observations without the cure fraction $\exp(-\theta)$. The number of subjects per group was chosen to be 100. The following configurations were considered: $\exp(-\theta) = 0.3, 0.5, 0.7, 0\%, 20\%$ and 40% censoring; $e^{\beta_1} = 2/3, 1, 3/2$ and $e^{\beta_2} = 0.5, 1, 2$. For 20% censoring as specified above, the actual rate of censoring was $44\%, 60\%$ and 76% for each plateau value. For 40% censoring, the actual rate of censoring was $58\%, 70\%$ and 82% for each plateau value. For each configuration, 1,000 replications were performed and the levels and powers of all tests were estimated at the nominal level of 0.05 .

4.2. Results

Tables 1(a–c) display the results for model (A) whereas Tables 2(a–c) display those for model (B).

Table 1(a) shows the results obtained in the uncensored case. Except for the LT test, the estimated level of each test under its proper null hypothesis is within the binomial range $[0.036, 0.064]$. In the presence of a short-term effect, the observed levels for the LT test increase up to 10% . The test of no short-term and long-term effect (SLT) shows a strongly increased power relative to LR, PPW and SLT-PH in the presence of a short-term effect. The power gains are striking for no, or small differences in the long-term effects. As compared to LR, the SLT test is in some cases 10 times more powerful. However, it is well known that the LR test is not suited for such situations where survival curves cross. In case of no difference in short-term effects, the power of this latter test is slightly decreased relative to that of the LR. In any case it is less than 12% lower than that of the LR test. Power values of the ST test are very close to the SLT. Regarding the long-term effect, the PL test is more powerful than LT and LR. Power of these two latter tests is quite close.

Table 1(b) shows the results obtained with a 20% censoring rate. The observed levels of the SLT and ST tests do not exceed the binomial bounds. This is not the case for LT and PL where the observed level is increased up to 9% in case a short-term effect exists. Concerning the power, it appears that the trends observed in the uncensored case remain almost unchanged. Power gains for the ST and SLT relative to LR are lower than in the uncensored case, but still remain impressive as compared to LR. For the LT, the magnitude of the power values is lower than in the uncensored case and is always under those of the PL test.

The results obtained at a 40% censoring rate are shown in table 1(c). Regarding the SLT and ST tests, empirical significance levels appear to be close to the nominal level, and power gains are less pronounced than at lower censoring rates. However, with a short-term effect, the magnitude of the power gain remains high. Concerning the LT test, the observed levels are appreciably higher than the nominal level in the presence of a short-term effect. For the PL test, the observed type I error rate is slightly increased up to 8% in case of an existing non-null short-term effect and power gains are less than those observed at a lower censoring rates.

Tables 2(a–c), show the results for model (B). Estimated type I error was very close to the nominal significance level of 0.05 for the SLT and the ST test in every configuration. This is not the case for the LT test which always yields higher observed levels than the nominal level with values markedly increased in some situations. To a lesser extent, a similar trend is observed for the PL test for which observed level is increased up to 18% in case of an existing non-null short-term effect and a high censoring rate. It is worth noting that in this short-term proportional hazards situation, the loss of power of the SLT test remains small as compared to the LR.

Concerning the type I error of the proposed tests, it should be stressed that the null hypotheses H_0 and H_{00} involve neither $\hat{\theta}$ nor $\hat{w}(t)$. As a result, the SLT and ST tests

maintain a correct type I error which is not the case for the LT test under the corresponding null hypothesis when β_2 is not null and the model is not the correct one. In the case of no short-term effect where the estimated level is close to the nominal one it appears that the power of the LT test is not dramatically decreasing as compared to the other tests even if in this case the uniform censoring is likely to hinder the long-term effect.

We performed additional simulations with small sample size (not shown here) and, as expected, it leads to a decrease in power which is more pronounced with a high censoring rate.

5. Application

In this section, we consider a clinical trial on breast cancer disease.

5.1. Primary Chemotherapy Trial

The aim of the present analysis was to investigate short-term and long-term effects of primary chemotherapy on disease recurrence by the proposed tests in a mature trial with more than ten years of follow-up. The so-called ‘S6-trial’ (Scholl et al., 1994) was conducted to assess whether primary chemotherapy improved survival, as compared to the same chemotherapy scheduled to follow the local regional treatment (adjuvant chemotherapy). Premenopausal breast cancer patients were included between October 1986 and June 1990, and randomized to receive either primary or adjuvant chemotherapy. The criteria for inclusion were as follows: non-metastatic operable breast tumors, largest tumor diameter between 3 and 7 cm, axillary lymph nodes not involved clinically, or involved but not adherent, no prior cancer, no serious concomitant illness. Bilateral, inflammatory or locally advanced breast cancers were not eligible. Two hundred breast cancer patients received primary chemotherapy and 190 adjuvant chemotherapy. Chemotherapy was started either after completion of the initial assessment (primary) or within 2 weeks of ending the local regional therapy (adjuvant). It consisted of four monthly cycles of intravenous cyclophosphamide, doxorubicin and 5-fluorouracil. Following random assignment to primary or adjuvant chemotherapy, patients were reviewed every 3 months for a year, then every 6 months during the first 5 years following the treatment and at least annually thereafter.

In what follows, we focus on the recurrence-free interval and not on the overall survival which was considered in a previous paper (Scholl et al., 1994). The recurrence-free interval is defined as the time from randomization until progression on the first observation of tumor recurrence (local, regional, distant).

5.2. Results

The median follow-up was 105 months. The 5-year recurrence-free interval rates were 60% [53–67] for patients treated with primary chemotherapy and 55% [48–63] for those treated with adjuvant chemotherapy. The 10-year survival rates were 40% [32–51] for patients treated with primary chemotherapy and 42% [35–50] for those treated with adjuvant chemotherapy. At the end of follow-up and for the 390 patients under study, 208 patients experienced a recurrence of the disease.

Figure 1 displays the Kaplan–Meier estimates of the recurrence-free interval by the treatment group. It shows a plateau value (i.e. long-term fraction) in the survival curves after 10 years, which is not surprising since most of the local and distant recurrences occur in the first decade (Bland and Copeland, 1998). Thus, an improper model appears well suited for these data.

Figure 2 displays the estimated survival function $A_i(t)$. The empirical estimate for $A_i(t) = [1 - (\Lambda_i(t)/\theta_i)]$ in each group is obtained by replacing $\Lambda_i(t)$ and θ_i by the Nelson–Aalen estimator and its value at the last observed failure time, respectively. This plot provides an informal assessment of the proportional hazards hypothesis for the short-term effect. It appears that the two survival functions cross, clearly indicating a non-proportional short-term effect.

In what follows, we present the results of the proposed statistics together with those obtained with the classical logrank statistic and the Peto–Prentice–Wilcoxon. We also provide the results of the SLT-PH test.

When testing for differences in recurrence-free interval, the logrank test ($\chi^2_1 < 0.0001, P=0.99$) the PPW ($\chi^2_1=0.10, P=0.78$) and the SLT-PH tests ($\chi^2_2=0.59, P=0.75$) are not significant. When testing for an overall effect with an accelerated failure time short-term effect, the SLT test is close to the significance ($\chi^2_2=4.14, P=0.13$). When testing for a short-term effect, the ST test is significant ($\chi^2_1=4.13, P=0.94$). No short-term effect was detected when using the LT test ($\chi^2_1 < 0.01, P=0.94$) or the plateau test ($PL: \chi^2_1=0.08, P=0.77$). These latter results agree with figure 2 which indicates a non-constant short-term effect. From these results, the diseases's recurrences have been significantly delayed by primary chemotherapy but without a benefit on long-term recurrence rate as compared to classical adjuvant chemotherapy.

6. Discussion

Survival data with long-term survivors requires extension of existing test statistics for analyzing short and long-term effects of a prognostic factor. In this paper, we propose new score tests well suited for different types of departures from equality of survival distributions with long-term survivors. These tests are related to improper short-term accelerated failure time alternatives and are obtained as score statistics from a time-dependent Cox model.

An interesting feature of these tests is that they are simple to use since they can be very easily obtained from standard Cox model procedures implemented in most statistical software packages. The test of no long-term effect should be particularized since its limiting distribution is obtained in the presence of a negligible short-term effect. This drawback would not exist if a test was derived from the original model, but this does not seem to be computationally realistic in usual practice. It must be kept in mind that using this test also requires that the maximum value of the cumulative hazard be estimated consistently. The theoretical condition underlying this assumption is that sufficient number of patients should be followed up to a time after which the risk of developing the event of interest is negligible. Such drawbacks do not concern the two other tests.

Simulation results show that SLT and ST tests maintain a correct type I error in case of high censoring rate and a misspecified model. In contrast, the proposed LT test is very sensitive to model misspecification and high censoring rates in the presence of a short-term effect. Regarding the power, the SLT and ST show interesting power performances for assessing a short-term effect with or without a long-term effect as compared to classical tests such as the logrank test or the Peto–Prentice–Wilcoxon test. Power gains decrease with censoring which could be explained by the fact that the cumulative hazard is not consistently estimated under the alternative hypothesis. Indeed, the presence of uniform censoring yields to a violation of the sufficient follow-up condition even if the model is correctly specified under the alternative hypothesis.

In practice, SLT and ST tests could be recommended for routine use when a non-constant short-term effect is expected. This could be the case when comparing treatments that modify the speed of progression of the disease in a population where a long-term survivor fraction is commonly encountered. As seen from the simulation results, it seems that with moderate censoring the product-limit test is a more reasonable alternative to the long-term effect tests when a short-term effect is expected.

The proposed score tests are particularly well suited to the study presented in this article since a large proportion of the patients will never recur from the disease and a long-term follow-up is provided. According to our analysis, the recurrence of the disease appeared to have been significantly delayed by primary chemotherapy but without a benefit on long-term recurrence rate as compared to the post-operative chemotherapy. It is tempting to speculate that early and effective targeting of active micrometastatic disease may have delayed the occurrence of disease recurrence. Based on these results, we should emphasize that using the proposed score tests provide some interesting findings for primary chemotherapy that would have been overlooked by only considering results from the classical logrank test. Moreover, we are able to attribute this difference to the short-term effect. Finally, our approach offers a new insight on the different aspects of treatment effects and may be recommended for widespread use in long-term survival studies.

In addition, it should be noted that these tests can obviously be extended for taking into account other factors by using a stratified time-dependent Cox model. Further works are ongoing to extend this family for taking into account other complex time-varying short-term patterns.

In conclusion, the proposed tests, which are very simple to use, seem particularly appealing when testing time-related effects of new markers or therapies in censored data with long-term survivors.

Acknowledgments

The authors would like to acknowledge Prof. A. Yakovlev (University of Utah) and Dr. B. Asselain (Institut Curie) for stimulating discussions during the evolution of this article. We thank all the members of the Breast Cancer Group of the Curie Institute for their cooperation. We are indebted to Mrs. C. Gautier for her kind technical assistance in carrying out the data management of the S6-trial. This work was supported in part by the 'Société Française du Cancer', NIH/NCI U01 CA 97414 and DAMD 17-03-1-0034.

Appendix. Partial Second Derivatives Based Upon the Working Model

For the following derivation it is convenient to introduce the notations:

$$\begin{aligned} S^{(0)}[\beta_1, \beta_2, \widehat{w}(t), t] &= \frac{1}{n} \sum_{k=1}^n Y_k(t) e^{\beta_1 Z_k} \exp\{Z_k \beta_2 [\widehat{w}(t)]\} \\ S^{(1)}[\beta_1, \beta_2, \widehat{w}(t), t] &= \frac{1}{n} \sum_{k=1}^n Z_k Y_k(t) e^{\beta_1 Z_k} \exp\{Z_k \beta_2 [\widehat{w}(t)]\} \\ S^{(2)}[\beta_1, \beta_2, \widehat{w}(t), t] &= \frac{1}{n} \sum_{k=1}^n Z_k^2 Y_k(t) e^{\beta_1 Z_k} \exp\{Z_k \beta_2 [\widehat{w}(t)]\} \end{aligned}$$

It follows that the partial second derivatives are as follows:

$$\begin{aligned}
\frac{\partial^2 \log L}{\partial \beta_1 \partial \beta_1} &= \sum_{j=1}^n \delta_j \left\{ \left\{ \frac{S^{(1)}[\beta_1, \beta_2, \widehat{w}(t_j), t_j]}{S^{(0)}[\beta_1, \beta_2, \widehat{w}(t), t_j]} \right\}^2 - \frac{S^{(2)}[\beta_1, \beta_2, \widehat{w}(t_j), t_j]}{S^{(0)}[\beta_1, \beta_2, \widehat{w}(t), t_j]} \right\} \\
\frac{\partial^2 \log L}{\partial \beta_2 \partial \beta_2} &= \sum_{j=1}^n \delta_j \left[\widehat{w}(t_j) \right]^2 \left\{ \left\{ \frac{S^{(1)}[\beta_1, \beta_2, \widehat{w}(t_j), t_j]}{S^{(0)}[\beta_1, \beta_2, \widehat{w}(t), t_j]} \right\}^2 - \frac{S^{(2)}[\beta_1, \beta_2, \widehat{w}(t_j), t_j]}{S^{(0)}[\beta_1, \beta_2, \widehat{w}(t), t_j]} \right\} \\
\frac{\partial^2 \log L}{\partial \beta_2 \partial \beta_2} &= \sum_{j=1}^n \delta_j \left[\widehat{w}(t_j) \right] \left\{ \left\{ \frac{S^{(1)}[\beta_1, \beta_2, \widehat{w}(t_j), t_j]}{S^{(0)}[\beta_1, \beta_2, \widehat{w}(t), t_j]} \right\}^2 - \frac{S^{(2)}[\beta_1, \beta_2, \widehat{w}(t_j), t_j]}{S^{(0)}[\beta_1, \beta_2, \widehat{w}(t), t_j]} \right\}
\end{aligned}$$

The elements of the information matrix are computed under the null hypothesis H_0 by using $\widehat{w}(t)$ as given in Section 3.1 and replacing $S^{(l)}(\beta_1, \beta_2, \widehat{w}(t), t)$ by $S^{(l)}(0, 0, \widehat{w}(t), t)$.

Under the (null hypothesis H_{00} , the corresponding elements are computed by using $\widehat{w}(t)$ as given in Section 3.1 and replacing $S^{(l)}(\beta_1, \beta_2, \widehat{w}(t), t)$ by $S^{(l)}(0, 0, \widehat{w}(t), t)$.

Finally, under the null hypothesis H_{000} , the corresponding elements are computed by using $\widehat{w}(t)$ as given in Section 3.1 and replacing $S^{(l)}(\beta_1, \beta_2, \widehat{w}(t), t)$ by $S^{(l)}(0, \widehat{\beta}_2, \widehat{w}(t), t)$.

References

- Aalen OO. Nonparametric inference for a family of counting processes. *Ann. Stat* 1978;6:701–726.
- Aalen OO. Modelling heterogeneity in survival analysis by the compound Poisson distribution. *Ann. Appl. Prob* 1992;2:951–972.
- Bland, KI.; Copeland, EM. *The Breast: Comprehensive Management of Benign and Malignant Diseases*. 2nd edn. Saunders; Philadelphia: 1998.
- Breslow NE. Contribution to the discussion on the paper by D.R. Cox' Regression and life tables. *J. Roy. Statist. Soc., B* 1972;34:216–217.
- Breslow NE. Covariance analysis of censored survival datas. *Biometrics* 1974;30:89–99. [PubMed: 4813387]
- Broët P, De Rycke Y, Tubert-Bitter P, Lellouch J, Asselain B, Moreau T. A semi-parametric approach for the two-sample comparison of survival times with long-term survivors. *Biometrics* 2001;57:844–852. [PubMed: 11550936]
- Cantor AB, Shuster JJ. Parametric versus non-parametric methods for estimating cure rates based on censored survival data. *Stat. Med* 1992;11:931–937. [PubMed: 1604072]
- Fleming, TR.; Harrington, DP. *Counting Processes and Survival Analysis*. John Wiley; New York: 1991.
- Gray RJ, Tsiatis AA. A linear rank test for use when the main interest is in differences in cure rates. *Biometrics* 1989;45:899–904. [PubMed: 2790128]
- Ibrahim JG, Chen MH, Sinha D. A new Bayesian model for survival data with a surviving fraction. *J. Amer. Statist. Assoc* 1999;94:909–919.
- Ibrahim JG, Chen MH, Sinha D. Bayesian semiparametric models for survival data with a cure fraction. *Biometrics* 2001;57:383–388. [PubMed: 11414560]
- Kabfleich, JD.; Prentice, RL. *The Statistical Analysis of Failure Time Data*. Wiley; New York: 1980.
- Kaplan E, Meier P. Nonparametric estimation from incomplete observations. *J. Am. Stat. Assoc* 1958;53:457–581.
- Kuk AYC, Chen C-W. A mixture model combining logistic regression with proportional hazards regression. *Biometrika* 1992;79:531–541.
- Laska EM, Meisner MJ. Nonparametric estimator and testing in a cure models. *Biometrics* 1992;48:1223–1234. [PubMed: 1290799]
- Lee JW. Two-sample rank tests for acceleration in cure models. *Stat. Med* 1995;14:2111–2118. [PubMed: 8552890]
- Maller, R.; Zhou, X. *Survival Analysis with Long-Term Survivors*. John Wiley; New York: 1996.

- Nelson W. Theory and applications of hazard plotting for censored failure data. *Technometrics* 1972;14:945–965.
- Peto R, Peto J. Asymptotically efficient rank invariant test procedures (with discussion). *J. Roy. Stat. Soc. Ser. A* 1972;135:185–198.
- Pierce DA. The asymptotic effect of substituting estimators for parameters in certain types of statistics. *Ann. Stat* 1982;10:475–478.
- Scholl SM, Fourquet A, Asselain B, Pierga J-Y, Vilcoq JR, Durand JC, Dorval T, Palangié T, Jouve M, Beuzeboc P, Garcio-Giralt E, Salmon RJ, De la Rochefordière A, Campana F, Pouillart P. Neoadjuvant versus adjuvant chemotherapy in premenopausal patients with tumours considered as too large for breast conserving surgery: Preliminary results of a randomised trial: S6. *Eur. J. Cancer* 1994;30A:645–652. [PubMed: 8080680]
- Spoto R, Sather HN, Baker SA. A comparison of tests of the difference in the proportion of patients who are cured. *Biometrics* 1992;48:87–99. [PubMed: 1581495]
- Taylor J. Semi-parametric estimation in failure time mixture models. *Biometrics* 1995;51:899–907. [PubMed: 7548707]
- Tsodikov A. A proportional hazards model taking account of long-term survivors. *Biometrics* 1998;54:1508–1516. [PubMed: 9883549]
- Tsodikov A. Semi-parametric models of long- and short-term survival: an application to the analysis of breast cancer survival in Utah by age and stage. *Stat. Med* 2002;30:895–920. [PubMed: 11870824]
- Yakovlev, A.; Tsodikov, A. *Stochastic Models of Tumor Latency and their Biostatistical Applications*. World Scientific; Singapore: 1996.

Figure 1.
Kaplan–Meier estimates of the recurrence-free interval according to the group of treatment.

Figure 2.
Estimated survival function $A(t)$ according to the group of treatment.

Table 1

Simulation results for the improper short-term accelerated failure time model with (a) no censoring, (b) 20% censoring and (c) 40% censoring.

$e^{-\theta}$	$e^{\beta 1}$	$e^{\beta 2}$	LR	PPW	SLT-PH	SLT	ST	LT	PL
<i>(a) No censoring</i>									
30%	0.67	0.5	0.482	0.219	0.362	0.993	0.995	0.478	0.598
		1.0	0.648	0.613	0.532	0.530	0.062	0.649	0.593
		2.0	0.757	0.873	0.657	0.999	0.992	0.787	0.575
	1.00	0.5	0.083	0.268	0.088	0.992	0.999	0.089	0.052
		1.0	0.043	0.046	0.048	0.048	0.056	0.040	0.046
		2.0	0.096	0.284	0.097	0.993	0.998	0.092	0.057
50%	1.50	0.5	0.901	0.988	0.857	1.000	1.000	0.858	0.626
		1.0	0.710	0.650	0.615	0.601	0.041	0.711	0.626
		2.0	0.411	0.099	0.407	1.000	1.000	0.557	0.657
	0.67	0.5	0.409	0.338	0.347	0.960	0.950	0.359	0.476
		1.0	0.466	0.445	0.379	0.366	0.056	0.470	0.472
		2.0	0.499	0.596	0.376	0.973	0.949	0.625	0.436
70%	1.00	0.5	0.052	0.070	0.050	0.957	0.978	0.073	0.049
		1.0	0.036	0.039	0.034	0.051	0.054	0.034	0.044
		2.0	0.049	0.088	0.054	0.950	0.972	0.094	0.048
	1.50	0.5	0.661	0.791	0.562	0.998	0.989	0.730	0.538
		1.0	0.575	0.563	0.471	0.476	0.063	0.579	0.569
		2.0	0.474	0.268	0.367	0.994	0.994	0.452	0.586
(b) 20% censoring	0.67	0.5	0.295	0.265	0.230	0.754	0.751	0.240	0.309
		1.0	0.314	0.308	0.231	0.241	0.057	0.321	0.317
		2.0	0.320	0.340	0.230	0.771	0.721	0.468	0.302
	1.00	0.5	0.055	0.056	0.044	0.719	0.831	0.093	0.048
		1.0	0.047	0.047	0.040	0.047	0.048	0.047	0.046
		2.0	0.049	0.053	0.048	0.748	0.839	0.081	0.051
(c) 40% censoring	0.67	0.5	0.440	0.479	0.348	0.925	0.884	0.571	0.401
		1.0	0.405	0.405	0.307	0.295	0.039	0.405	0.391
		2.0	0.398	0.335	0.299	0.913	0.906	0.332	0.435
	1.00	0.5	0.055	0.056	0.044	0.719	0.831	0.093	0.048
		1.0	0.047	0.047	0.040	0.047	0.048	0.047	0.046
		2.0	0.049	0.053	0.048	0.748	0.839	0.081	0.051
(d) 60% censoring	0.67	0.5	0.440	0.479	0.348	0.925	0.884	0.571	0.401
		1.0	0.405	0.405	0.307	0.295	0.039	0.405	0.391
		2.0	0.398	0.335	0.299	0.913	0.906	0.332	0.435
	1.00	0.5	0.055	0.056	0.044	0.719	0.831	0.093	0.048
		1.0	0.047	0.047	0.040	0.047	0.048	0.047	0.046
		2.0	0.049	0.053	0.048	0.748	0.839	0.081	0.051

$e^{-\theta}$	e^{β_1}	e^{β_2}	LR	PPW	SLT-PH	SLT	ST	LT	PL
		0.5	0.518	0.234	0.683	0.980	0.984	0.530	0.615
	0.67	1.0	0.570	0.548	0.464	0.467	0.057	0.574	0.460
		2.0	0.722	0.869	0.663	0.997	0.990	0.737	0.499
		0.5	0.079	0.322	0.683	0.972	0.979	0.084	0.106
30%	1.00	1.0	0.053	0.051	0.063	0.055	0.057	0.054	0.058
		2.0	0.102	0.297	0.371	0.981	0.991	0.075	0.053
		0.5	0.908	0.988	0.986	1.000	0.986	0.845	0.204
	1.50	1.0	0.569	0.523	0.479	0.422	0.051	0.563	0.348
		2.0	0.184	0.039	0.730	0.957	0.991	0.451	0.492
		0.5	0.438	0.305	0.434	0.905	0.902	0.377	0.475
	0.67	1.0	0.424	0.416	0.348	0.342	0.049	0.432	0.429
		2.0	0.500	0.574	0.393	0.949	0.904	0.618	0.407
		0.5	0.049	0.095	0.415	0.850	0.927	0.082	0.087
50%	1.00	1.0	0.048	0.049	0.063	0.056	0.063	0.049	0.049
		2.0	0.060	0.086	0.142	0.860	0.933	0.079	0.061
		0.5	0.680	0.820	0.863	0.973	0.908	0.703	0.162
	1.50	1.0	0.405	0.382	0.315	0.305	0.059	0.400	0.298
		2.0	0.298	0.128	0.577	0.890	0.948	0.398	0.475
		0.5	0.326	0.293	0.261	0.693	0.662	0.269	0.311
	0.67	1.0	0.275	0.274	0.217	0.202	0.047	0.287	0.279
		2.0	0.287	0.318	0.211	0.716	0.661	0.421	0.275
		0.5	0.045	0.038	0.192	0.544	0.674	0.055	0.071
70%	1.00	1.0	0.053	0.057	0.055	0.057	0.052	0.057	0.057
		2.0	0.044	0.055	0.096	0.560	0.658	0.071	0.046
		0.5	0.342	0.414	0.560	0.767	0.649	0.436	0.138
	1.50	1.0	0.274	0.279	0.214	0.213	0.050	0.286	0.228
		2.0	0.244	0.183	0.375	0.632	0.734	0.276	0.342
<i>(c) 40% censoring</i>									
		0.5	0.198	0.084	0.707	0.892	0.954	0.270	0.474
	0.67	1.0	0.474	0.457	0.393	0.379	0.047	0.475	0.318
		2.0	0.704	0.869	0.868	0.987	0.940	0.610	0.158

$e^{-\theta}$	e^{β_1}	e^{β_2}	LR	PPW	SLT-PH	SLT	ST	LT	PL
30%	1.00	0.5	0.303	0.535	0.790	0.948	0.950	0.283	0.056
		1.0	0.050	0.054	0.061	0.046	0.050	0.054	0.050
		2.0	0.197	0.447	0.727	0.948	0.966	0.110	0.069
	1.50	0.5	0.976	0.992	0.991	1.000	0.911	0.947	0.407
		1.0	0.500	0.473	0.394	0.379	0.055	0.485	0.295
		2.0	0.052	0.097	0.719	0.911	0.984	0.185	0.322
50%	0.67	0.5	0.243	0.160	0.557	0.747	0.839	0.249	0.434
		1.0	0.327	0.316	0.239	0.236	0.042	0.328	0.246
		2.0	0.425	0.553	0.555	0.880	0.816	0.466	0.173
	1.00	0.5	0.093	0.150	0.514	0.772	0.825	0.132	0.044
		1.0	0.054	0.059	0.052	0.047	0.050	0.061	0.053
		2.0	0.083	0.144	0.475	0.770	0.843	0.083	0.081
70%	1.50	0.5	0.807	0.890	0.923	0.973	0.749	0.817	0.277
		1.0	0.341	0.328	0.259	0.257	0.056	0.336	0.230
		2.0	0.091	0.057	0.552	0.734	0.896	0.207	0.354
	0.67	0.5	0.229	0.187	0.378	0.503	0.582	0.229	0.307
		1.0	0.213	0.217	0.161	0.172	0.054	0.224	0.189
		2.0	0.213	0.243	0.230	0.538	0.473	0.282	0.130
	1.00	0.5	0.046	0.056	0.287	0.443	0.574	0.072	0.054
		1.0	0.040	0.040	0.047	0.037	0.044	0.043	0.044
		2.0	0.050	0.058	0.231	0.468	0.596	0.069	0.062
	1.50	0.5	0.504	0.557	0.643	0.773	0.468	0.563	0.191
		1.0	0.229	0.228	0.150	0.157	0.045	0.236	0.172
		2.0	0.120	0.085	0.382	0.472	0.661	0.193	0.293

Configurations presented below correspond to different short-term effect (β_1), long-term effect (β_2), and plateau values ($e^{-\theta}$). The total number of subjects is 200.

Table 2

Simulation results for the improper short-term proportional hazard model with (a) no censoring, (b) 20% censoring and (c) 40% censoring.

e^{-0}	$e^{\beta 1}$	$e^{\beta 2}$	LR	PPW	SLT-PH	SLT	ST	LT	PL
<i>(a) No censoring</i>									
30%	0.67	0.5	0.887	0.963	0.989	0.823	0.888	0.882	0.601
		1.0	0.609	0.591	0.517	0.519	0.048	0.611	0.579
		2.0	0.303	0.111	0.880	0.220	0.895	0.311	0.568
	1.00	0.5	0.207	0.450	0.870	0.186	0.894	0.196	0.046
		1.0	0.050	0.048	0.046	0.050	0.041	0.051	0.061
		2.0	0.209	0.487	0.896	0.193	0.916	0.186	0.039
50%	1.50	0.5	0.208	0.057	0.889	0.150	0.909	0.225	0.600
		1.0	0.711	0.674	0.600	0.592	0.038	0.692	0.609
		2.0	0.972	0.994	0.997	0.956	0.873	0.963	0.646
	0.67	0.5	0.624	0.758	0.922	0.503	0.812	0.624	0.463
		1.0	0.480	0.475	0.393	0.381	0.047	0.480	0.487
		2.0	0.310	0.188	0.800	0.221	0.810	0.314	0.444
70%	1.00	0.5	0.084	0.192	0.783	0.062	0.854	0.084	0.046
		1.0	0.048	0.048	0.043	0.056	0.052	0.048	0.051
		2.0	0.087	0.188	0.784	0.065	0.853	0.085	0.051
	1.50	0.5	0.310	0.146	0.878	0.224	0.882	0.314	0.573
		1.0	0.579	0.574	0.490	0.488	0.061	0.581	0.569
		2.0	0.847	0.935	0.978	0.739	0.866	0.840	0.567
(b) 20% censoring	0.67	0.5	0.353	0.418	0.686	0.253	0.606	0.354	0.293
		1.0	0.322	0.327	0.252	0.239	0.047	0.328	0.317
		2.0	0.257	0.200	0.570	0.176	0.543	0.255	0.311
	1.00	0.5	0.053	0.073	0.540	0.043	0.660	0.053	0.054
		1.0	0.047	0.047	0.047	0.051	0.042	0.048	0.053
		2.0	0.066	0.086	0.562	0.047	0.681	0.067	0.059
(c) 40% censoring	1.50	0.5	0.291	0.206	0.763	0.216	0.713	0.292	0.402
		1.0	0.405	0.394	0.300	0.305	0.044	0.409	0.398
		2.0	0.505	0.602	0.841	0.381	0.739	0.507	0.387

$e^{-\theta}$	e^{β_1}	e^{β_2}	LR	PPW	SLT-PH	SLT	ST	LT	PL
		0.5	0.928	0.971	0.980	0.909	0.686	0.920	0.496
	0.67	1.0	0.559	0.534	0.456	0.447	0.049	0.555	0.476
		2.0	0.146	0.055	0.677	0.153	0.720	0.161	0.463
		0.5	0.404	0.561	0.623	0.476	0.462	0.358	0.069
30%	1.00	1.0	0.042	0.038	0.044	0.040	0.053	0.035	0.046
		2.0	0.438	0.584	0.636	0.482	0.461	0.391	0.058
		0.5	0.039	0.066	0.207	0.084	0.286	0.041	0.106
	1.50	1.0	0.563	0.531	0.466	0.429	0.049	0.546	0.314
		2.0	0.989	0.994	0.987	0.986	0.238	0.981	0.590
		0.5	0.699	0.787	0.891	0.615	0.677	0.706	0.407
	0.67	1.0	0.421	0.403	0.319	0.315	0.050	0.423	0.398
		2.0	0.207	0.130	0.652	0.165	0.665	0.207	0.416
		0.5	0.236	0.316	0.458	0.283	0.420	0.232	0.055
50%	1.00	1.0	0.049	0.050	0.050	0.045	0.050	0.047	0.052
		2.0	0.205	0.288	0.467	0.249	0.447	0.205	0.057
		0.5	0.058	0.057	0.188	0.079	0.235	0.066	0.125
	1.50	1.0	0.429	0.416	0.343	0.329	0.051	0.431	0.297
		2.0	0.932	0.946	0.916	0.909	0.234	0.918	0.557
		0.5	0.405	0.460	0.654	0.317	0.529	0.421	0.286
	0.67	1.0	0.296	0.290	0.225	0.226	0.043	0.304	0.293
		2.0	0.213	0.179	0.470	0.170	0.458	0.207	0.296
		0.5	0.106	0.133	0.289	0.138	0.326	0.115	0.050
70%	1.00	1.0	0.052	0.052	0.050	0.045	0.044	0.056	0.054
		2.0	0.111	0.138	0.299	0.146	0.321	0.126	0.067
		0.5	0.054	0.048	0.152	0.086	0.183	0.059	0.095
	1.50	1.0	0.289	0.286	0.213	0.217	0.042	0.296	0.229
		2.0	0.738	0.768	0.698	0.673	0.185	0.740	0.430
(c) 40% censoring									
		0.5	0.959	0.971	0.946	0.943	0.221	0.948	0.587
	0.67	1.0	0.448	0.421	0.345	0.335	0.036	0.456	0.292
		2.0	0.052	0.049	0.169	0.079	0.222	0.051	0.079

$e^{-\theta}$	e^{β_1}	e^{β_2}	LR	PPW	SLT-PH	SLT	ST	LT	PL
30%	1.00	0.5	0.573	0.624	0.527	0.507	0.143	0.538	0.209
		1.0	0.049	0.050	0.060	0.048	0.051	0.055	0.053
		2.0	0.559	0.618	0.523	0.525	0.141	0.527	0.184
	1.50	0.5	0.083	0.085	0.100	0.087	0.094	0.070	0.043
		1.0	0.447	0.427	0.359	0.338	0.050	0.440	0.244
		2.0	0.990	0.988	0.976	0.972	0.086	0.976	0.700
50%	0.67	0.5	0.819	0.847	0.786	0.763	0.203	0.815	0.463
		1.0	0.326	0.328	0.257	0.246	0.053	0.332	0.247
		2.0	0.077	0.063	0.177	0.073	0.219	0.083	0.125
	1.00	0.5	0.364	0.408	0.345	0.342	0.142	0.351	0.143
		1.0	0.047	0.039	0.039	0.034	0.044	0.050	0.036
		2.0	0.373	0.410	0.368	0.342	0.143	0.358	0.147
70%	1.50	0.5	0.057	0.064	0.097	0.067	0.108	0.052	0.051
		1.0	0.330	0.319	0.269	0.254	0.060	0.329	0.231
		2.0	0.919	0.928	0.883	0.873	0.097	0.911	0.632
	0.67	0.5	0.547	0.572	0.505	0.469	0.172	0.555	0.333
		1.0	0.213	0.212	0.171	0.169	0.042	0.225	0.201
		2.0	0.058	0.049	0.125	0.076	0.184	0.063	0.122
	1.00	0.5	0.178	0.208	0.219	0.193	0.146	0.193	0.114
		1.0	0.062	0.058	0.042	0.051	0.044	0.066	0.055
		2.0	0.225	0.239	0.208	0.185	0.127	0.231	0.111
	1.50	0.5	0.064	0.063	0.071	0.064	0.074	0.064	0.059
		1.0	0.206	0.211	0.143	0.164	0.056	0.206	0.158
		2.0	0.669	0.679	0.591	0.596	0.079	0.666	0.459

Configurations presented below correspond to different short-term effect (β_1), long-term effect (β_2), and plateau values ($e^{-\theta}$). The total number of subjects is 200.