Paper 125-2010

Using SAS® to Assess and Model Time-to-Event Data with Non-Proportional Hazards

Michael G. Wilson
Indianapolis IN, USA

ABSTRACT

Proportional Hazards Regression using a partial maximum likelihood function to estimate the covariate parameters (Cox, 1972) has become an exceedingly popular procedure for conducting survival analysis. It is a notably robust survival method because it makes no assumptions about the shape of the probability distribution for survival times. On the other hand, strong violations of the proportional hazards assumption can have detrimental effects on the validity and efficiency of the partial likelihood inference (Struthers and Kalbfleisch, 1986; Lin and Wei, 1989). In this paper, graphical and analytical procedures to assess violations and extensions of the model to improve inferential validity and efficiency in the presence of non-proportionality using SAS® are compared and presented.

1. Introduction

Most statistical methods for the analysis of time-to-event data can be classified based on the distributional assumption as non-parametric, semi-parametric and parametric. Generically, the name for this time is *survival time*, although it may be applied to time 'survived' to any event of interest.

Two features make special methods called *survival analysis* necessary. Firstly, the event of interest is observed in only a subset of individuals and subsequently survival times are not observed in a subset. This phenomenon is called censoring. Secondly, these data are rarely normally distributed, but are skewed.

SAS/STAT® procedures LIFETEST, PHREG and LIFEREG, respectively provide a comprehensive set of tools to draw valid and reliable conclusions from these complex data.

Non-parametric methods are available in the LIFETEST procedure when the adjustment for only a few covariates is necessary. The procedure provides survival probabilities for constructing the survivorship function, S(t) (Kaplan and Meier, 1985). In addition, survival in two or more groups can be compared using the log-rank test (Peto elt al, 1977).

LIFEREG can be used to fit Accelerated failure time (AFT) models using maximum likelihood methods. AFT models describes the relationship between the survivor functions, S(t) for two groups.

$$S(t) = S_0(\varphi t)$$

The acceleration factor is φ and will stretch or shrink the survival curve along the time axis by a constant relative amount. If the acceleration factor is less than one, ($\varphi < 1$), the length of survival is decreased compared with the baseline survivor function. Conversely, if the acceleration factor is greater than one ($\varphi > 1$), the length of survival is increased.

In the four decades since its introduction, the proportional hazards regression (Cox 1972) has been established as the first choice of many persons wanting to perform regression analysis of censored survival data (Bradburn et al 2003). PHREG has emerged as a powerful SAS procedure to conduct such analyses. Its capabilities can be greatly extended by use of a variety of public domain macros as well as customization techniques. Practical applications occur not only in medical research but also in economics, industrial reliability and the agricultural, biological and physical sciences.

The proportional hazards (PH) regression model has two kinds of assumptions, that when satisfied allows one to rely on the statistical inferences and predictions the model yields. The first assumption is that the relationship between log hazard or log cumulative hazard and a covariate is linear. The second assumption is the time independence of the covariates in the hazard function, that is, the ratio of the hazard function for two individuals with different regression covariates does not vary with time, which is also known as the PH assumption.

The application of a statistical method to data in which the model assumptions are violated can result in wrong conclusions. Although there are several approaches to detecting, testing and modeling non-proportional hazards in the literature, few researchers propose methods for focusing on verifying assumptions at the onset of the analysis.

In order to understand how to test the proportional hazards assumption, it is important to understand the hazard function.

2. The Hazard Function

The hazard rate is the number of events which occur during a unit of time. National average hazard rates for one year for selected events are shown in Table 1.

Table 1. Annual Risks for Selected Events

Event	Annual Risk
Car Stolen	1 in 100
House fire	1 in 200
Die of Heart Disease	1 in 280
Win State Lottery	1 in 1 million
Killed by flood or tornado	1 in 2 million
Die in commercial plane crash	1 in 10 million

Source: National Aeronautics and Space Administration, 2009
http://www.cotf.edu/ete/modules/volcanoes/vrisk.html

For example, Hu and colleagues reported data from 86,016 women who were followed for 1,132,229 person years and classified them by nut consumption. He and his team reported that women who were nut consumers had 30 fatal heart attacks in 100,000 person years (hazard rate = 0.00030) compared to 49 fatal heart attacks per 100,000 person years (hazard rate = 0.00049) for those women who did not consume nuts Hu *et. al.* 1998).

Importantly however, the chance of a dying of heart disease is different for an 87 year old than a 10 year old. So when the hazard rate is described over time, it is called the hazard rate as a function of time, $h(t)$, or hazard function for short.

Figure A illustrates several generic shapes of hazard functions. For short periods of time, a hazard function can be constant (shown in black). A hazard function for an event can be increasing with age (red). A hazard function could be decreasing as in the case of patients recovering from kidney transplants (blue). The bathtub shape can describe a life-time hazard function (orange). If the hazard rate increases early and eventually declines then the function is arch shaped (green) and this is often used to model survival after successful surgery where early there is an increase risk of infection, bleeding or other complication followed by a decline in risk as the patient recovers.

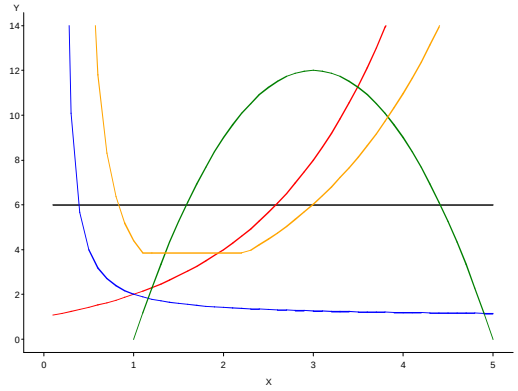


Figure A. Shapes of hazard functions: constant (black); increasing (red); decreasing (blue); bathtub shaped (orange); arch-shaped (green)

The hazard function is usually more informative about the underlying mechanism of failure than the survival function. Because of its convenience, the hazard function is more popular way of describing continuous survival data than the probability density function (pdf). The hazard function is a limiting function of time that quantifies the instantaneous risk an event will occur at time t and is formally defined as:

$$h(t) = \lim_{\Delta t \to 0} \frac{\Pr\{t \le T < t+\Delta t | T \ge t\}}{\Delta t} \qquad (2.1)$$

The hazard function is the limit of the instantaneous conditional probability an event occurs within the interval between t and Δt. Firstly, because time is continuous, the probability that an event will occur at exactly an instantaneous time t is zero, so we take the limit as the interval between t and Δt goes to zero. Secondly, we divide this probability by Δt. Thirdly, for a given interval, the probability is conditional on surviving to time t, because those who died before time t are not at risk for the event.

The equation in (2.1) is not a probability because values can be greater than one. The equation can be simplified in terms of the probability density function and the survivor function:

$$h(t) = \frac{f(t)}{S(t)}, \qquad (2.2)$$

where f(t) is the probability density function (pdf) and S(t) is the survivor function. So there is a special relationship among f(t), F(t), S(t) and h(t), such that if given one the other three can be derived.

Three important and well-known hazard functions can be derived based on how their logarithm is related to time. If the logarithm of the hazard is constant in time then the failure times have an exponential distribution (see

appendix). When the log hazard is linear in time the model follows a Gompertz distribution. When the log hazard is linear in the log of time, the model follows a Weibull distribution function.

Each of the logarithm models can be extended to account for the influence of covariates, which are used as explanatory variables. For example, age might be one covariate that explains an increasing hazard. The log hazards are provided in Table 2 for the case where there are p covariates $[x_1, x_2, \ldots, x_p] = \mathbf{X}^t$.

Table 2. Log Hazards for Three Distributional Models

Distribution	$\log h(t)$	
Exponential	$\mu +$	$[\beta_1 x_1 + \ldots + \beta_p x_p]$
Gompertz	$\mu + \alpha t +$	$[\beta_1 x_1 + \ldots + \beta_p x_p]$
Weibull	$\mu + \alpha \log(t) +$	$[\beta_1 x_1 + \ldots + \beta_p x_p]$

The summation in the square brackets of Table 2 can be rewritten as below in the spirit of the usual linear models formulation for the effects of covariates.

$$\left[\beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p\right] = \sum_{i=1}^{p} \beta_i x_i \quad (2.3)$$

The coding of factors and their interaction effects follows the usual rules for linear models. For example, if a factor has four levels, three indicator variables may be constructed to model the effect of the factor. An interaction between two or more factors may be examined by constructing new variables which are the product of the variables associated with the individual factors as is usually done in linear models. One needs to take care in interpreting coefficients so constructed.

When one hazard function with covariates $\mathbf{X}^t$ is divided by another with covariates $\mathbf{X}^{t*}$, the quotient is called the hazard ratio. The hazard ratio is the relative risk of an individual with risk factors $\mathbf{X}^t$ having the event compared to an individual with risk factors $\mathbf{X}^{t*}$.

When the hazard ratio is one, the risk of the event is equal. If the hazard ratio is greater than one, the factors increase the change of the event precariously. Conversely, if the hazard ratio is less than one, the factors are protective. For example, in the nut consumption study women who eat nuts had a lower risk of having heart attacks because the relative risk was 0.61 (0.00030/0.00049; Hu *et. al.* 1998).

Admittedly the hazard ratio is perhaps a little non-intuitive and in survival analysis doesn't describe how much longer an individual will live. In the future, statistics like the restricted mean could take its place (Schaubel and Zhang 2010). Until then, we will be using the hazard function and hazard ratios.

When the hazard ratio of functions with different covariates is constant, the hazards are said to be proportional.

$$h_1(t) = \gamma\, h_2(t) \quad (2.4)$$

Otherwise, the gamma term above is a function of time and the hazard functions are said to be non-proportional.

$$h_1(t) = \gamma(t)\, h_2(t). \quad (2.5)$$

3. Proportional Hazards Regression

Like many other models, the PH regression models the hazard function, as can be seen in equation 4.1.

$$h(t) = h_0(t)\, exp\left\{\sum_{i=1}^{p} \beta_i x_i\right\} \quad (3.1)$$

In the PH model, the hazard function is dependent on, or determined by, a set of p covariates $[x_1, x_2, \ldots, x_p]$, whose impact is measured by the size of the respective coefficients $[\beta_1, \beta_2, \ldots, \beta_p]$. The 't' in $h(t)$ reminds us that the hazard function varies over time. The term $h_0(t)$ is called the baseline hazard, and is the value of the hazard if all the x_i are equal to zero, since the quantity exp(0) equals 1. The proportional hazards model is considered semi-parametric because no assumption regarding the distribution of the baseline hazard is necessary.

The quantities $\exp(\beta_i)$ are called hazard ratios. A value of β_i greater than zero, or equivalently a hazard ratio greater than one, indicates that as the value of the i^{th} covariate increases, the event hazard increases and thus the length of survival decreases. In other words, a hazard ratio above one indicates the covariate is positively associated with the event probability, and thus negatively associated with the length of survival.

The PH model is essentially a multiple linear regression of logarithm of the hazard on the variables, x_i, with the baseline hazard being an 'intercept' term that varies with time. The covariates then act multiplicatively on the hazard at any point in time, and this provides us with the key assumption of the PH model: the hazard of the event in any group is a constant multiple of the hazard in any other. This assumption implies that the hazards for groups should be proportional and cannot cross or diverge. This

proportionality assumption is often appropriate for survival time data, but in some cases where it is inappropriate can lead to false conclusions.

The process of modeling proportional hazards is admittedly fluid and iterative so many assumptions should not be ignored. Wilson (2008) recommended a thorough checking for at least five potential problems and provided some recommendations on possible solutions.

Table 3. Possible Remedial Measures for Issues in PH Modeling

	Potential Problem	Remedial Measure
1.	Outliers and Influential Observations	PH Influence Statistics
2.	Interaction among Covariates	Martingale Plots
3.	Incorrect Functional Form for the Covariates	Cumulative Residual Plots
4.	Quasi-complete separation	Firths Penalized Likelihood Maximum
5.	Correlated Responses	Quasi-likelihood estimation of the Sandwich Variance

In this paper, we restrict ourselves to the problem of the proportional hazards violation and assume the potential problems listed above have been adequately addressed.

4. Methods

Synthetic, censored time-to-event data for six pre-defined hazard patterns were generated using a clinical trial format for acute coronary syndrome. Specifically, three patient variables were generated including a randomly assigned dichotomous treatment variable, continuous age (in years) and diagnostic cardiac troponin-T (cTnT; in µg/L) variable (Melanson 2007). The continuous variables were independent, centered, standardized and from a normal distribution, with coefficients of -1, 2 and 0 for the age-by-cTnT interaction.

The six patterns generated are illustrated in Figure B and included two adhering to the proportional hazards (PH) assumption since the ratio of hazard functions is constant: (1) the null hypothesis case and (2) the alternative hypothesis case. The four remaining patterns were departures from the PH assumption since the ratio of hazard functions was not constant including (3) decreasing, (4) increasing, (5) diverging and (6) crossing hazards.

The failure times were generated from the Weibull hazard, $h(t) = \lambda\gamma(\lambda t)^{\gamma-1}$ for each of the six patterns using the parameter values from Table 4 (Bender 2005). The censoring mechanism for all patterns was singly, fixed (Type I) at 5 years.

Table 4. Weibull Parameter Values for Six Hazard Patterns

		Control Group		Investigational Group	
Case	Hazards Pattern	Shape (λ)	Scale (γ)	Shape (λ)	Scale (γ)
1	Constant (1)	2.00	2.00	2.00	2.00
2	Constant (2)	1.00	1.00	1.00	2.00
3	Decreasing	0.30	2.00	0.50	2.00
4	Increasing	1.50	2.00	2.00	2.00
5	Diverging	0.75	1.00	2.00	1.00
6	Crossing	1.50	1.00	3.00	1.00

Adapted from Ng'andu 1997

Histograms for the synthetic failure times are provided in Figure C which also shows the kernel for 5 probability distributions including the normal, lognormal, exponential, Weibull and Gamma. Further, summary statistics are provided in the inset.

All failure times are non-negative and their distribution is right skewed. As a result, the sample means are expected to be larger than the theoretically expected values. The theoretically expected value in the absence of censoring is given by $\Gamma(1+1/\gamma)/\lambda^{1/\gamma}$. For example, in the null hypothesis case the observed mean is 0.865 and the theoretically expected value in the absence of censoring is 0.627 (SAS Data Step shown in Table 5).

Table 5. SAS Data Step for the Calculation of the Theoretically Expected Value for Weibull = $\Gamma(1+1/\gamma)/\lambda^{1/\gamma}$

```
*** nph04s01.sas ***;
data;
 shape = 2; scale = 2;
 mean = gamma(1+(1/shape))/shape**(1/scale);
 put mean; run;
```

Routinely we also create plots of the survival curves (Figure D), the hazard function (Figure E) and the cumulative hazard (Figure F) using the LIFETEST procedure.

5. Graphical Checks

Since the validity of inferences based on the PH model depends on the proportional hazards assumption, it is desirable to have diagnostic methods for checking this assumption. Many tools are available for checking the PH assumption. These include plots of (a) Log Cumulative Hazard, (b) Schoenfeld Partial Residuals, and (c) Standardized Score Process.

5a. Log Cumulative Hazard

With the additional constraint that the hazard functions are proportional, the baseline hazard function algebraically cancels out of the numerator and denominator. Specification of the underlying distribution is unnecessary making the model semi-parametric. The remaining parameters can be estimated using partial maximum likelihood estimation.

Suppose the hazards for two groups are proportional, as in equation 2.4 above: $h_1(t) = \gamma\, h_2(t)$. It can be shown that this results in the relationship:

$$\log\left[-\log S_1(t)\right] = log\gamma + \log\left[-\log S_2(t)\right] \qquad (5.1)$$

This relationship implies that a plot of the log(-log S(t)) curves for the two groups would differ by a constant. We recognize -log S(t) as the cumulative hazard function, H(t). So plotting the estimated log(-log S(t)) curves for the two groups, provides a visual check of the PH assumption. A clear departure from parallelism of these two curves would be consistent with violation of the PH assumption.

The survival probabilities used in the construction of this plot can be unadjusted and come from the non-parametric estimation, which can be obtained in LIFETEST using the lls option. On the other hand, the survival probabilities can be adjusted for covariates and estimated from the semi-parametric survivorship curve, which are obtained from PHREG using the baseline option. In the adjusted case, the log cumulative hazard functions are evaluated at the covariate mean values. In practice, both are produced to get different perspectives from the data, but are often similar. If one were to produce only a single form, the unadjusted represents the least susceptible to contamination.

The Unadjusted Log Cumulative Hazard plots by Log time for the Six Synthetic Cases are shown in Figure G. The null case shows the similarity and shape of the curves. In the Alternative case the curves are separated by the treatment effect. The treatment effect in the Decreasing, Increasing and Diverging cases is separated early but merges toward the end of the study. As expected, the treatment effect crosses in the Crossing case.

The Adjusted Log Cumulative Hazard by Log time plots for the Six Synthetic Cases are shown in Figure H. These plots require an extra Data Step and a GPLOT procedure in SAS 9.1. The violation of PH in the Crossing case is clearer in the unadjusted case, but otherwise, the results are similar to the unadjusted plots.

The Adjusted Log Cumulative Hazard by the original time scales for the Six Synthetic Cases are shown in Figure I. Under the PH condition, these curves would be expected to differ by a constant value, but not be parallel. Therefore, these plots would be considered less sensitive and it would be more difficult to detect violations.

5b. Schoenfeld Partial Residuals

For each covariate in a PH regression, a Schoenfeld residual can be calculated for each case that was not censored. Under the proportional hazards assumption, a plot of these residuals against time should be "approximately flat" (Grambsch and Therneau [10]). These residuals are available using the RESSCH option on the OUTPUT statement in PHREG. To make it easier to detect violations of the PH assumption, some authors recommend superimposing a LOESS line and looking for a non-zero trend.

Figure J shows the Schoenfeld residual plot for a continuous covariate known to follow the proportional hazards assumption. As expected, the distribution of Schoenfeld residuals are evenly distributed about zero. The density of residuals lightens at longer survival times due to the decrease in sample size.

On the other hand, Figure K shows the Schoenfeld residual plot for the categorical treatment variable for the six synthetic cases. Under the proportional hazards assumption of the first two cases, plots of these residuals against time are approximately flat. However the violation is easily detected in the last two cases.

5c. Standardized Score Process

The standardized empirical score process is a transform of the martingale residuals. A martingale is a special sequence of random variables where the conditional expected value of the next observation, given all the past observations, is equal to the last observation.

These standardized score paths are a 'Tied Down Brownian Process' since they start and end a zero. Several processes or paths can be simulated under the null hypothesis and plotted with the observed path. If the observed path is typical of the simulated paths it is considered evidence of proportional hazards. Atypical observed paths are evidence of violation of proportional hazards.

Figure L provides the standardized score processes for the Alternative and Crossing cases. While there is support for the PH assumption for the Constant Hazards – Alternative0Hypothesis case for all three covariates, there

appears to be a violation of PH in the Crossing cases for treatment.

6. Tests for Non-Proportionality

Graphical checks are useful but subjective. So they can be augmented with analytical tests. Several tests have been suggested including, tests based on (a) on re-sampling, (b) on Schoenfeld Residuals, (c) the Hazard Ratio for the covariate by time interaction, and (d) on the Hazard Ratios for the interaction between the covariates and categorized time.

6(a) Standardized Score Process Test

Therneau et al. (1990) proposed testing the PH assumption using the score process. This PH statistic is sensitive to alternatives for which covariates have a monotonically increasing or decreasing effect over time. This test statistic has no known distribution; however, Lin *et al*. (1993) have shown it is consistent, against non-proportional hazards alternatives. In SAS®, this test is labeled 'Supremum Test for Proportionals Hazards Assumption'

For the Crossing Hazards case, Table 6 shows the Maximum Absolute Value and the *P*-value for the Supremum Test for Proportionals Hazards Assumption. The Maximum Absolute Values for the covariates are less than 1.96, indicating no evidence of the violation of proportional hazards. On the other hand, the maximum absolute value for the treatment covariate is greater than 1.96, which is statistically significant at the 0.05 alpha level as shown by a significant *P*-value.

6(b) Schoenfeld Partial Residuals Test

Harrell (1986) developed a computationally simple test of the PH assumption based on Schoenfeld's partial residuals of the model. It is based on Fisher's *z*-transform of the Pearson correlation between the partial residuals and the rank order of the failure time.

These residuals do not depend on time and they do not involve an estimated baseline hazard function which simplifies their asymptotic distribution. When there are tied failure times, one takes the residual as the total component of the first derivatives of the log-likelihood function with respect to a regression parameter divided by the number of tied failure times at the corresponding risk set, and weights the correlation estimate by the number of tied times.

The test statistic for testing $\rho = 0$, that is, that PH holds, is a normal deviate calculated by the formula:

$$Z = \rho\sqrt{(n_u - 2)/(1-\rho^2)}, \qquad (6.1)$$

where ρ is the correlation between residuals and failure time order and n_u is the total number of uncensored observations. The test statistic tends to be positive if the ratio of the hazards for high values of the covariate increases over time, and it tends to be negative if this hazard ratio decreases over time. It requires no categorization of the time variable or the covariate.

This statistics can easily be calculated in a Data Step and is not available from the PHREG procedure.

6(c) Hazard Ratio for the covariate by time interaction

In his original paper, Cox (1972) proposed a way to check the PH assumption by introducing a constructed time-dependent variable, that is, add to the model interaction terms that involve time (for example, treatment–by–log (*t*)) and test for their significance. Importantly, the partial likelihood function has the same form with and without these time-dependent covariates.

Therefore, to check the PH assumption for time-dependent variables, fit an extended Cox model that contains time-dependent variables defined with some function of time:

$$h(t;\boldsymbol{X}(t)) = h_0(t)\, exp\left\{\sum_{i=1}^{p} \beta_i \mathbf{X}_i + \sum_{i=1}^{p} \gamma_i \mathbf{X}_i g_i(t)\right\} \qquad (6.2)$$

where $g_i(t)$ is some nonzero function of time that corresponds to $\mathbf{X}_i$ (for example, $g_i(t) = \log(t)$ or $g_i(t) = \text{rank}(t)$). The hazard ratio is a constant for all *t* only when $\gamma_i = 0$.

To test the null hypothesis that $\gamma = 0$, that is, whether PH is adequate, one can compute the likelihood ratio test statistic using:

$$-2\ln\left[\frac{L(\hat{\beta},0)}{L(\hat{\beta},\hat{\gamma})}\right] \sim X^2, \qquad (6.3)$$

with appropriate degrees of freedom.

The creation of the interaction with time is complex data manipulation because that value changes. The PHREG

procedure is exceptional for creating these variables because it provides a rich subset of DATA step operators and functions for defining time-dependent covariates. These operators follow the model statement in PHREG as illustrated in Table 7.

Table 7. SAS Code for Testing PH with the Interaction between Treatment and Time

```
*** nph06s01 ***;
proc phreg data = ads01;
     model survtime*event(0) =
           x1 x2 x3 trt trttime
           / ties = efron;
      trttime = trt*logsurvtime;
     run;
```

When the covariate is known to satisfy the assumption of proportional hazards, as age in the case of the Crossing Hazards, the likelihood chi-square is 1.7974 (Table 8) and is not significant. On the other hand, the treatment covariate has a likelihood chi-square of 2,549.0946 and is significant (Table 9).

6(d) Hazard Ratios for the interaction between the covariates and categorized time.

Some programmers might be concerned that the interaction effect is not linear. So instead of dichotomizing, time could be categorized into four or five groups. Within these groups, the assumption that the covariate effect is linear is more reasonable. These interactions can be tested using the same likelihood test. Similar to the case of interaction between continuous time and the covariate, the partial likelihood function performs well. If the hazard ratios from these interactions are similar, they can be dropped.

Table 10. SAS Code for Testing PH with the Interaction between Treatment and Categorized Time

```
*** nph06s01 ***;
proc rank data = ads01
          out = ads04 group = 5;
     var survtime;
     ranks rsurvtime;
     run;

proc phreg data = ads04;
     model survtime*event(0) = x1 x2 x3
           trttime0 trttime1 trttime2
           trttime3 trttime4
           / ties = efron;
     trttime0 = trt * rsurvtime_0;
     trttime1 = trt * rsurvtime_1;
     trttime2 = trt * rsurvtime_2;
     trttime3 = trt * rsurvtime_3;
     trttime4 = trt * rsurvtime_4;
     run;
```

The choice for the number of intervals should be based on subject-matter knowledge. Furthermore, there should be a relatively equal number of events and censored observation across the time intervals to ensure that the standard errors of the parameter estimates are relatively similar.

The Data Step to create the time categories is illustrated in Table 10. Care in creating these indicator variables should be exercised. Because of missing values, Boolean programming is not recommended.

The results in Table 11 show that over the five time categorized groups, the hazard ratio is increasing and therefore, provides evidence that the PH assumption is violated.

6(e) Comparison of PH Test Performance

Ng'andu (1997) showed that these three test statistics, that is the Score Process test, the Schoenfeld Partial Residuals Test and the Hazard Ratio for the covariate by time interaction are practically equally powerful. The Interaction Test for Continuous time has the advantage of its simplicity.

7. Modeling Non-Proportionality

7(a) Modeling the Time by Covariate Interaction

When the proportional hazards assumption is violated, the effect of the predictor variable varies with time. Cox (1972) proposed an obvious solution when he suggested the inclusion of a new variable in the model from Section 6(c). From that section, it is known that this new variable is the interaction between the predictor variable and continuous time.

Also, as was shown, the values of this new variable changes with over time. These variables differ from time-independent variables where the values were determined at baseline (time = 0) and these values did not change over the period of observation.

7(c) Stratification

The PHREG procedure allows one to model the non-proportionality by stratification. This technique is most useful when the covariate that interacts with time is categorical, not of direct interest and too difficult to model. Stratification is limited by the fact that one cannot stratify by a variable and also include it as a covariate.

The basic idea of the stratified PH model is that the baseline hazard function is allowed to vary across strata. In other words, the underlying hazard function for on strata

Table 6. Partial output from PHREG using the ASSESS option

Supremum Test for Proportionals Hazards Assumption

Variable	Maximum Absolute Value	Replications	Seed	Pr > MaxAbsVal
trt	5.3397	1000	46163	<.0001
x1	1.2089	1000	46163	0.1530
x2	1.2693	1000	46163	0.1750

Table 8. Maximum Likelihood Estimates from PHREG procedure for Interaction between Age (X1) and Time

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio
x1	1	-0.80131	0.01273	3962.2248	<.0001	0.449
x2	1	1.61956	0.01625	9935.5621	<.0001	5.051
x3	1	0.01907	0.01099	3.0130	0.0826	1.019
trt	1	0.11559	0.02313	24.9682	<.0001	1.123
x1time	1	-0.00327	0.00244	1.7974	0.1800	0.997

Table 9. Maximum Likelihood Estimates from PHREG procedure for Interaction between Treatment (TRT) and Continuous Time

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio
x1	1	-0.90746	0.01236	5388.3594	<.0001	0.404
x2	1	1.80893	0.01735	10868.2407	<.0001	6.104
x3	1	0.08778	0.01018	74.3177	<.0001	1.092
trt	1	-0.54525	0.02627	430.7541	<.0001	0.580
trttime	1	0.60965	0.01208	2549.0946	<.0001	1.840

Table 11. Output from PHREG procedure for Interaction between Treatment (TRT) and Categorical Time

The PHREG Procedure

Model Information

Data Set	WORK.ADS04
Dependent Variable	survtime
Censoring Variable	event
Censoring Value(s)	0
Ties Handling	EFRON

Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Without Covariates	With Covariates
-2 LOG L	164217.86	150925.09
AIC	164217.86	150941.09
SBC	164217.86	150998.77

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	13292.7637	8	<.0001
Score	13477.7631	8	<.0001
Wald	11173.4468	8	<.0001

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio
x1	1	-0.85988	0.01307	4327.4703	<.0001	0.423
x2	1	1.72533	0.01952	7811.5800	<.0001	5.614
x3	1	0.02681	0.01075	6.2132	0.0127	1.027
trttime0	1	-0.21911	0.06389	11.7617	0.0006	0.803
trttime1	1	-0.11454	0.04248	7.2682	0.0070	0.892
trttime2	1	-0.03552	0.03583	0.9824	0.3216	0.965
trttime3	1	0.10308	0.03430	9.0299	0.0027	1.109
trttime4	1	0.58883	0.04436	176.1789	<.0001	1.802

can be completely different from the underlying hazard function for other.

A stratified PH model ranks the event times separately within strata. However, a common vector of regression coefficients is fitted across the strata. These parameter estimates can be thought of as pooled estimates.

So a stratified PH model can be used to obtain a separate underlying survival curve for each stratum while adjusting of the other predictor variables that satisfy the proportional hazards assumption. In fact, the model can be used to assess the proportional hazards assumption by plotting the predicted survival curves from a model without stratification and the predicted survival curves from a model with stratification. Major differences indicate that the assumption is violated (Kleinbaum 1996).

One of the main disadvantages of stratification is that no parameter estimate and no hazard ratio are obtained for the stratification variable. Stratified PH models are used when the stratification variables are known to affect the outcome but the estimates of the effects are considered to be of secondary importance (Hosmer and Lemeshow 1999)

8. Conclusion

These results show that fitting a proportional hazards (PH) model to non-proportional hazards data can lead to incorrect conclusions. Good graphical and analytical methods for detecting violations of the PH assumption were identified and their implementation demonstrated in SAS®. Three strategies for properly modeling non-proportional time-to-event data were provided and their advantages and disadvantages were discussed.

9. References

Allison, Paul D. (1995), Survival Analysis Using the SAS System: A Practical Guide, Cary, NC: SAS Institute Inc.

Bender R, Augustin T, Blettner M (2005). Generating survival times to simulate Cox proportional hazards models. Statist. Med. 2005; 24:1713-1723.

Bradburn MJ, Clark TG, Love SB, Altman DG (2003) Br J Cancer 89, 431 – 436.

Cox, D. (1972), Regression Models and Lifetables (with discussion), Journal of the Royal Statisticial Society, Series B, 34, 187-220.

Melanson SEF, Tanasijevic M, Jarolim P. Cardiac Troponin Assays. Circulation. 2007; 116:e501-e504.

Schaubel DE and Zhang M (2010). Semiparametric Estimator for Differences in Restricted Mean Lifetimes in Observational Studies. Joint Statistical Meetings, Section 441,Vancouver, B.C. August 2010.

Hosmer, D.W., Jr., and Lemeshow, S. (1999), Applied Survival Analysis: Regression Modeling of Time to Event Data. New York: Wiley.

Hu FB, Stampfer MJ, Manson JE, Rimm EB, Colditz GA, Rosner BA, Speizer FE, Hennekens CH, Willett WC. BMJ 1998; 317; 1341-1345.

Kaplan EL, Meier P (1958) Nonparametric estimation from incomplete observations. J Am Stat Assoc 53: 457-481.

Kleinbaum, D.G. (1996), Survival Analysis: A Self-Learning Text, New York: Springer-Verlag.

Le, Chap T. (1997), Applied Survival Analysis. New York: Wiley.

Lee, ET. (1992), Statistical Methods for Survival Data Analysis, Second Edition, New York: John Wiley & Sons, Inc.

Lin, D. Y. and Wei, L. J. (1989). The robust inference for the Cox proportional hazards model. J. Am. Statist. Assoc. 84, 1074-8.

Ng'andu NH. (1997) An Empirical Comparison of Statistical Tests for Assessing the Proportional Hazards Assumption of Cox's Model. Statis. Med. 1997;16:611-626.

Peto R, Pke MC, Armitage P, Breslow NE, Cox DR. Howard SV, Mantel N, McPherson K, Peto J, Smith PG (1977) Design and analysis of randomized clinical trials required prolonged observation of each patient. II. Analysis and examples. Br J Cancer 35:1-39.

Pocock S, Clayton TC, Altman DG (2002) Survival plots of time-to-event outcomes in clinical trials: good practice and pitfalls. Lancet 359: 1686 – 1689.

Struthers, C. A. and Kalbfleisch, J. D. (1986). Misspecified proportional hazard models. Biometrika 73, 363-9.

Wilson MG. Introduction to Predictive Analytics Using the SAS® System. Annual Meeting of the Iowa SAS® Users Group, Two-hour, Invited, May 15th, 2009, Des Moines, Iowa.

Wilson MG. Introduction to Survival Analysis Using SAS®. Annual Meeting of the Mid-West SAS® Users Group, Three-hour, Invited In-Conference Training, October 12-14th, 2008, Indianapolis IN.

10. Recommended Reading

Altman, D.G. Schulz, K. F., Moher, D. et. al. (2001), The Revised CONSORT Statements for Reporting Randomized Trails: Explanation and Elaboration, Annals of Internal Medicine, 134, 663-694.

Barrett-Connor, E. (1991), Postmenopausal Estrogen and Prevention Bias, Annals of Internal Medicine, 115, 455-456.

Cox, D. (1972), Regression Models and Lifetables (with discussion), Journal of the Royal Statisticial Society, Series B, 34, 187-220.

Cox, D. R. and Oakes, D. (1984), analysis of Survival Data, London: Chapman & Hall.

Fleming TR and Harrington DP. (2005), Counting Processes and Survival Analysis, New York: Wiley.

Freedman, David (2008). Survival Analysis: A Primer. The American Statistician, 62(2), 110-119.

Halley, E. (1693), An Estimate of the Mortality of Mankind, Drawn from Curious Tables of the Births and Funerals at the City of Breslaw; with an Attempt to Ascertain the Price of Annuities upon Lives, Philosophical

Transactions of the Royal Society of London, 196,596-610, 654-656.

Harrell, F. E. The PHGLM procedure, SAS supplemental Library ºseries Guide, Version 5 Edition, SAS Institute, Cary, N.C., 1986.

Kalbfleish, J.D. and Prentice, R.L. (2002), The Statistical Analysis of Failure Time Data, 2nd Edition, New York: John Wiley & Sons, Inc.

Kaplan, E. L., and Meier, P. (1958), Nonparametric Estimation from Incomplete Observations, Journal of American Statistical Association, 53, 457-481.

Klein, J.P. and Moeschberge, M.L. (1997), Survival Analysis: Techniques for Censored and Truncated Data, New York: Springer-Verlag.

Lawless, J. F. (2003), Statistical Models and Methods for Lifetime Data (2nd ed.), New York: Wiley.

Lin, D. Y., Wei, L. J. and Ying, Z. Checking the Cox model with cumulative sums of martingale-based residuals, Biometrika, 80, 557-572 (1993).

Miller, Rupert G. (1981), Survival Analysis, New York: John Wiley &Sons, Inc.

Pargament, K. I., Koenig, H. G., Tarakeshwar, N., and Hahn, J (2001), Religious Struggle as a Predictor of Mortality among Medically Ill Patients, Archives of Internal Medicine, 161, 1881-1885.

Therneau, T. M., Grambsch, P. M. and Fleming, T. R. Martingale-based residuals for survival models, Biometrika, 77, 147-160 (1990).

Wei, L. J. Testing goodness of fit for proportional hazards model with censored observations, Journal of the American Statistical Association, 79, 649-652 (1984).

ACKNOWLEDGMENTS

The author thanks the MWSUG 2010 co-chairs and Statistical Section Chairman for selecting this contribution for presentation. In addition, he is grateful for the reviewers who made suggestions for improvements and discovered errors. Nevertheless, any remaining errors are my full responsibility.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact me at:

Name:	Michael G. Wilson
Address:	1252 S. Broken Arrow, Dr.
City, State ZIP:	New Palestine, IN 46163
Work Phone:	317.861.1947
Fax:	317.861.5226
E-mail:	micgwils@iupui.edu

TRADEMARK INFORMATION

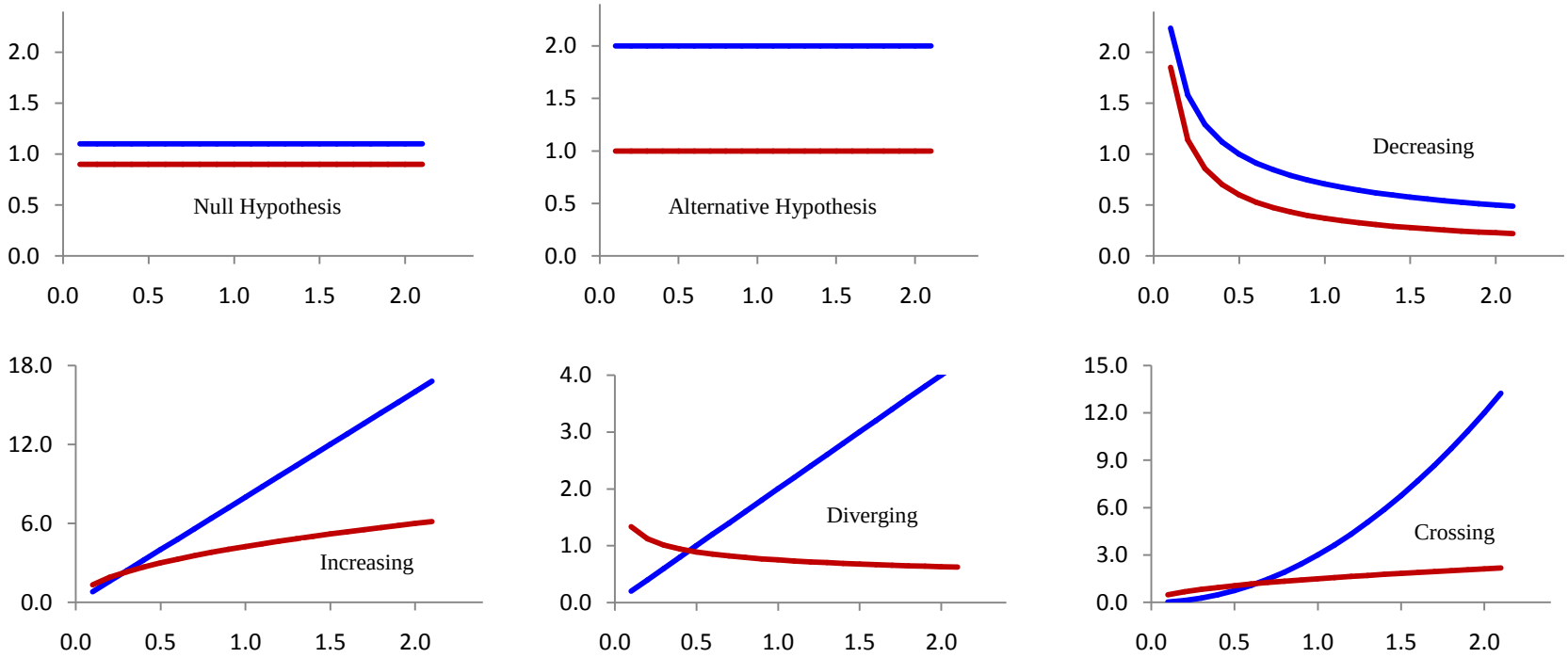
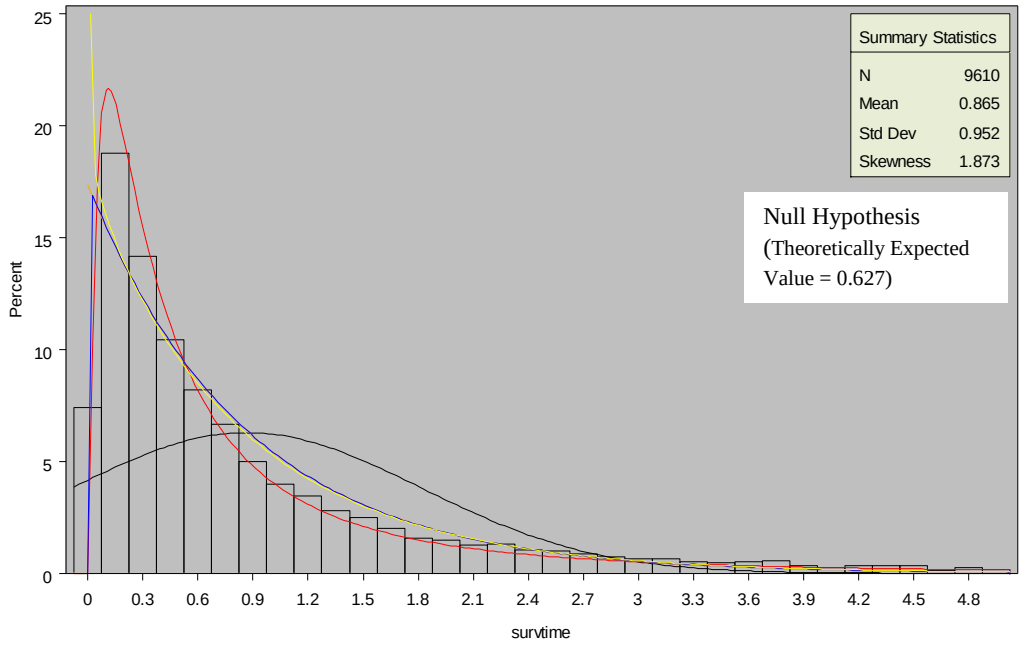
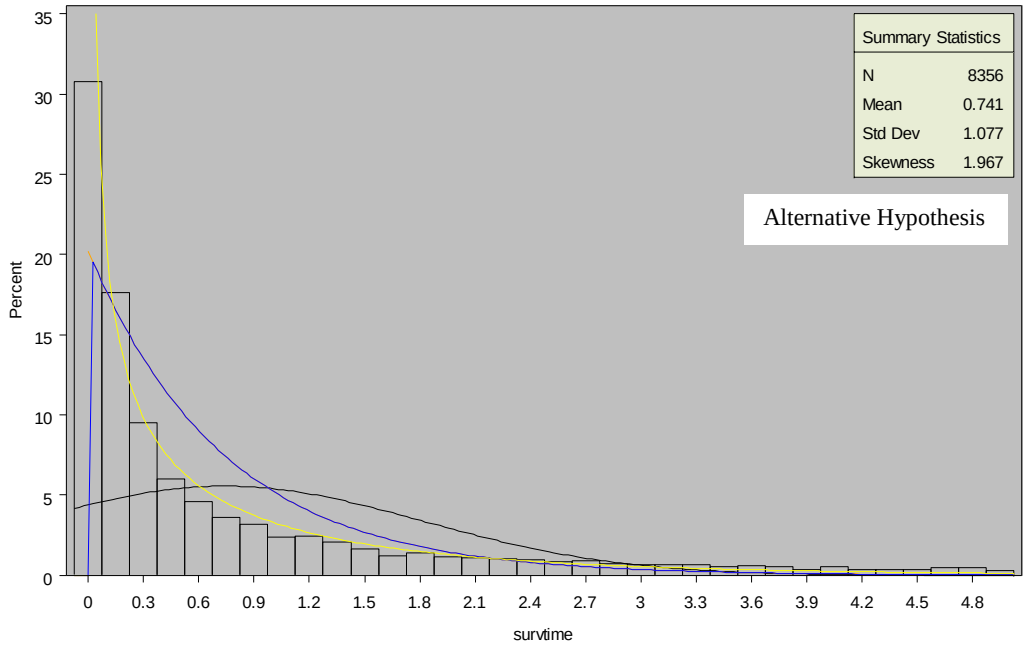


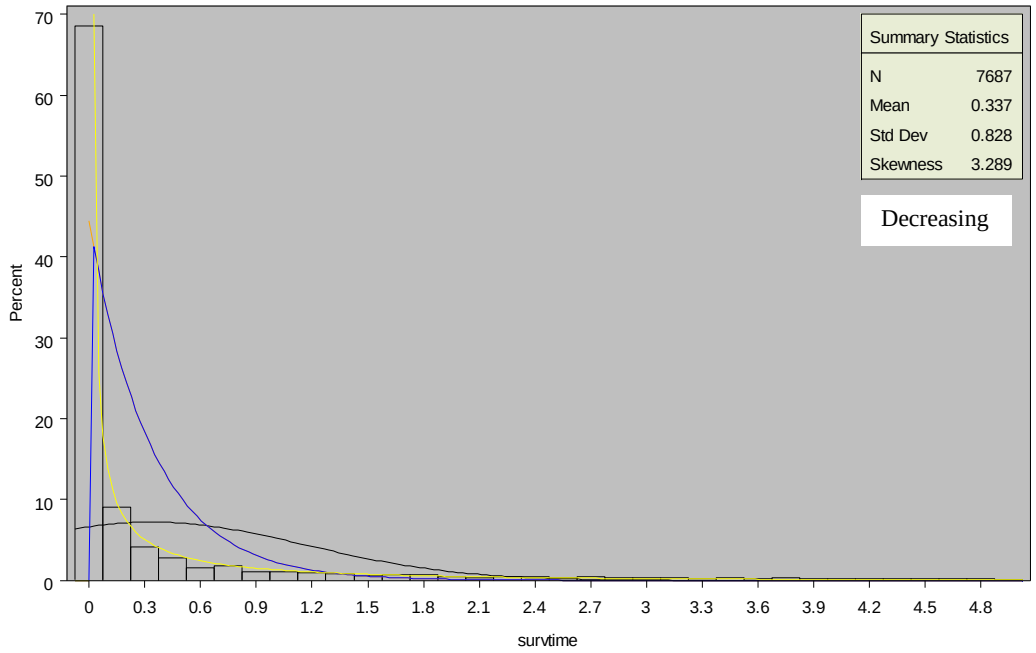
Figure B. Six Patterns of Hazard Functions by Time. The control groups are shown in red while investigational groups are shown in blue. Top Left Panel (Constant – Null); Top Middle (Constant – Alternative); Top Right (Decreasing); Bottom Left (Increasing); Bottom Middle (Diverging); Bottom Right (Crossing).



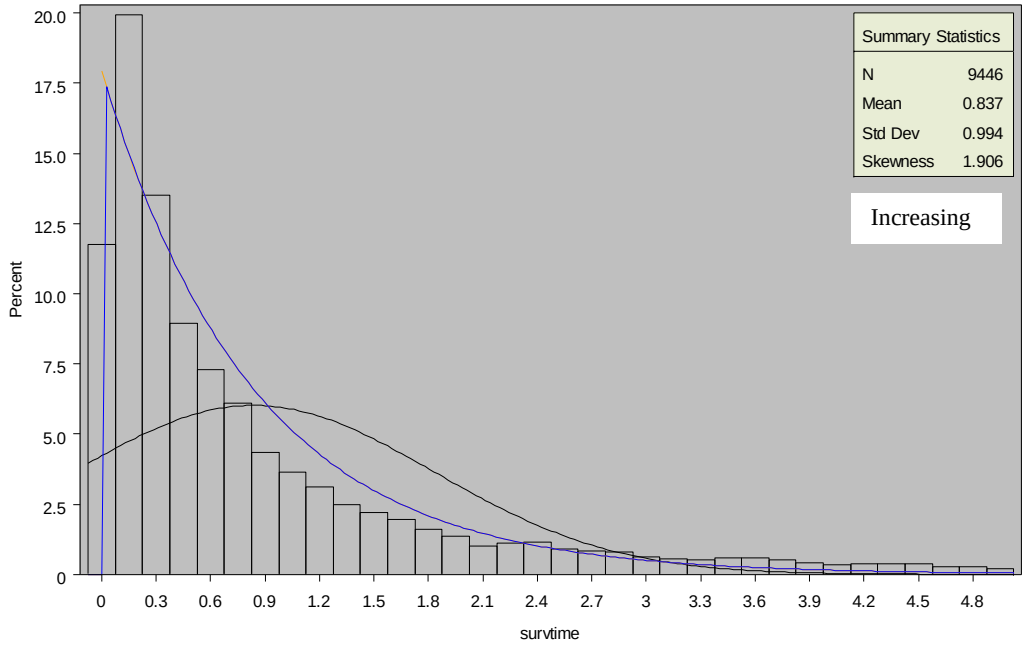
Normal = Black, LogNormal = Red, Exponential = Orange, Weibull = Blue, Gamma = Yellow



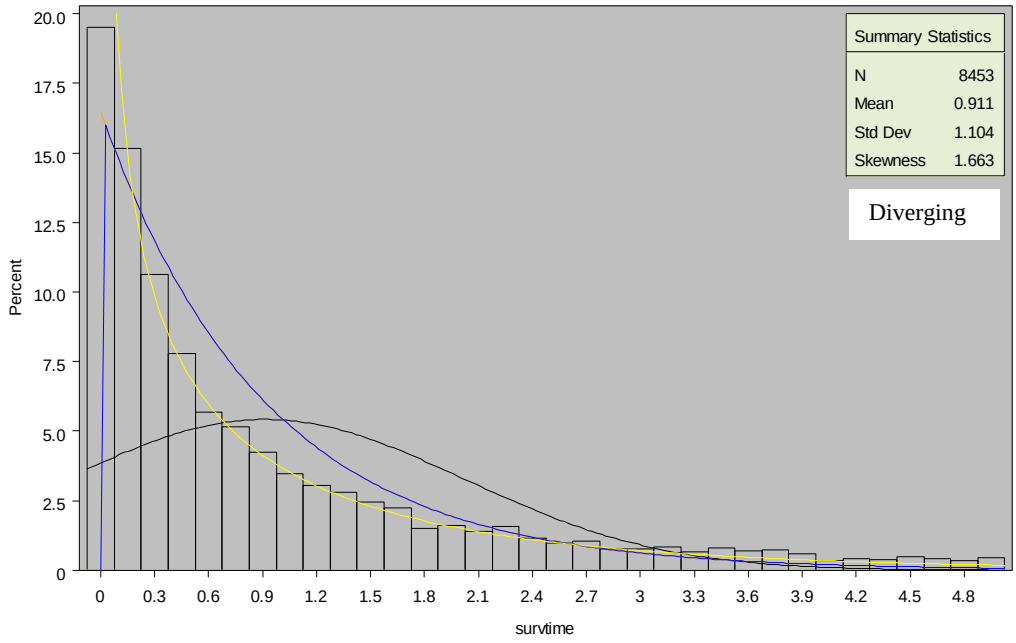
Normal = Black, LogNormal = Red, Exponential = Orange, Weibull = Blue, Gamma = Yellow



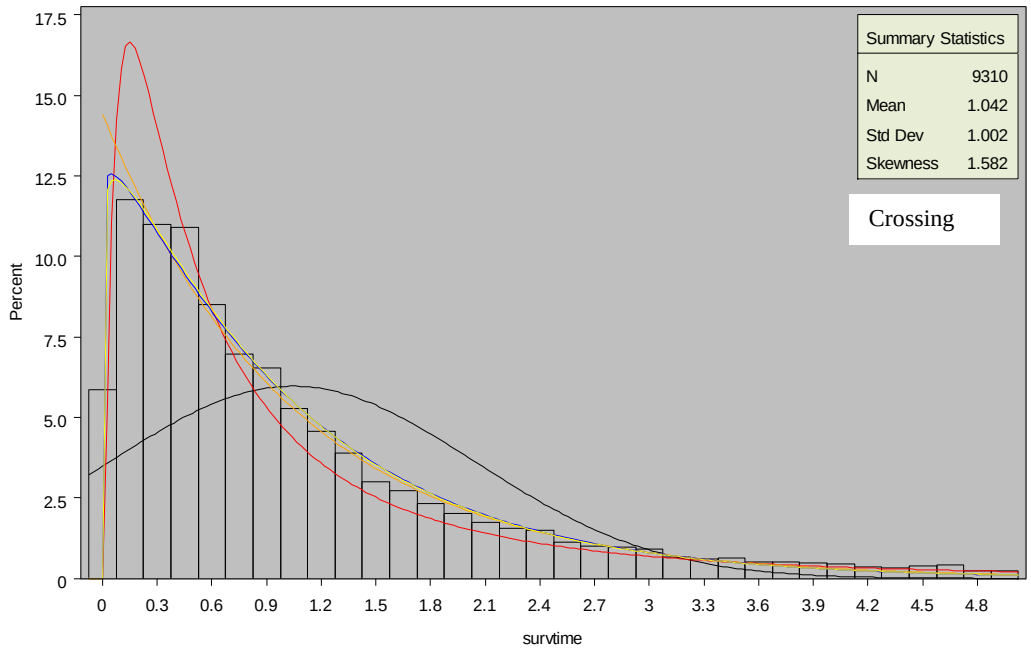
Normal = Black, LogNormal = Red, Exponential = Orange, Weibull = Blue, Gamma = Yellow



Normal = Black, LogNormal = Red, Exponential = Orange, Weibull = Blue, Gamma = Yellow



Normal = Black, LogNormal = Red, Exponential = Orange, Weibull = Blue, Gamma = Yellow



Normal = Black, LogNormal = Red, Exponential = Orange, Weibull = Blue, Gamma = Yellow

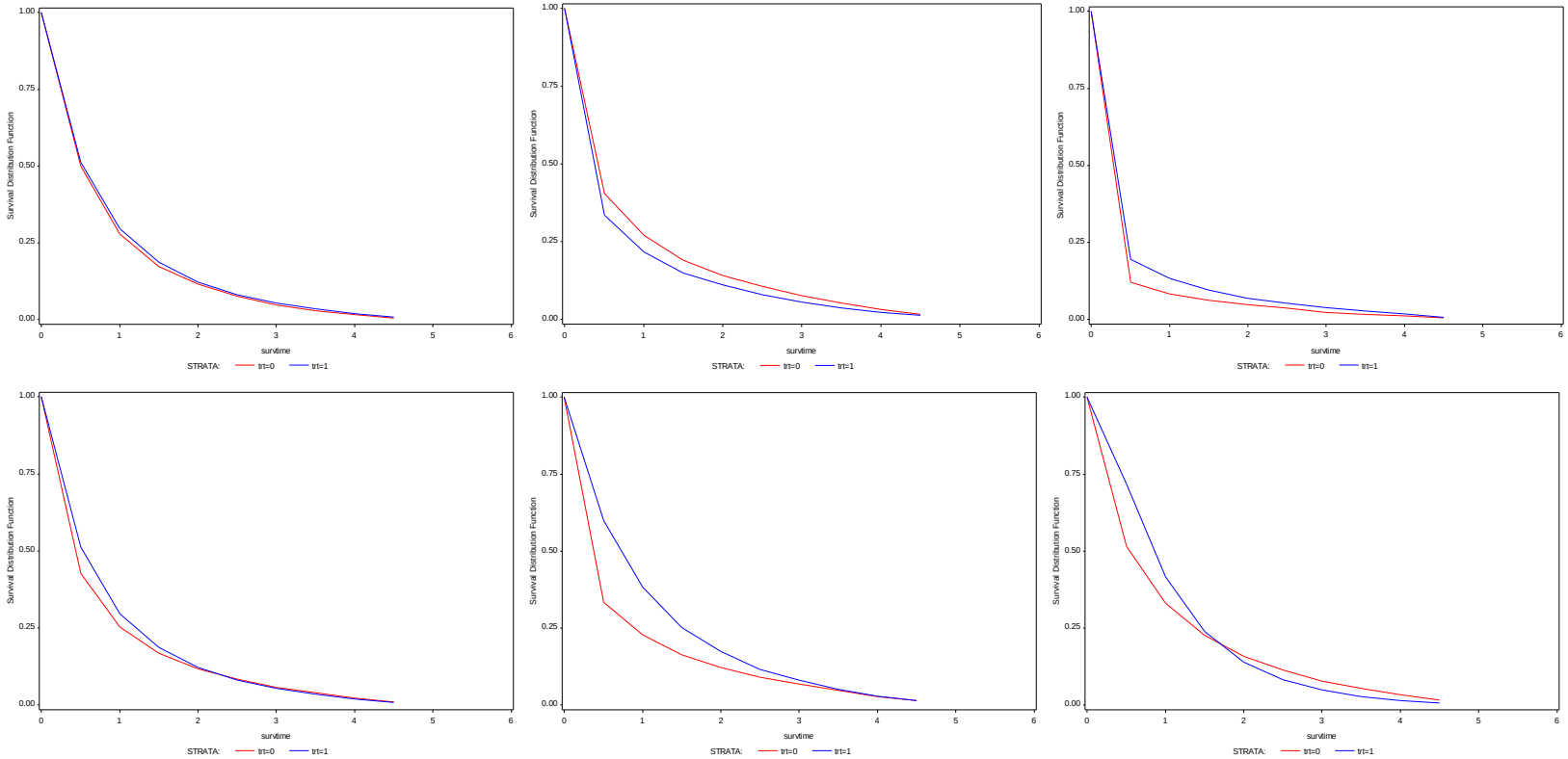


Figure D. Six Patterns of Survival Functions by Time. The control groups are shown in red while investigational groups are shown in blue. Top Left Panel (Constant – Null); Top Middle (Constant – Alternative); Top Right (Decreasing); Bottom Left (Increasing); Bottom Middle (Diverging); Bottom Right (Crossing).

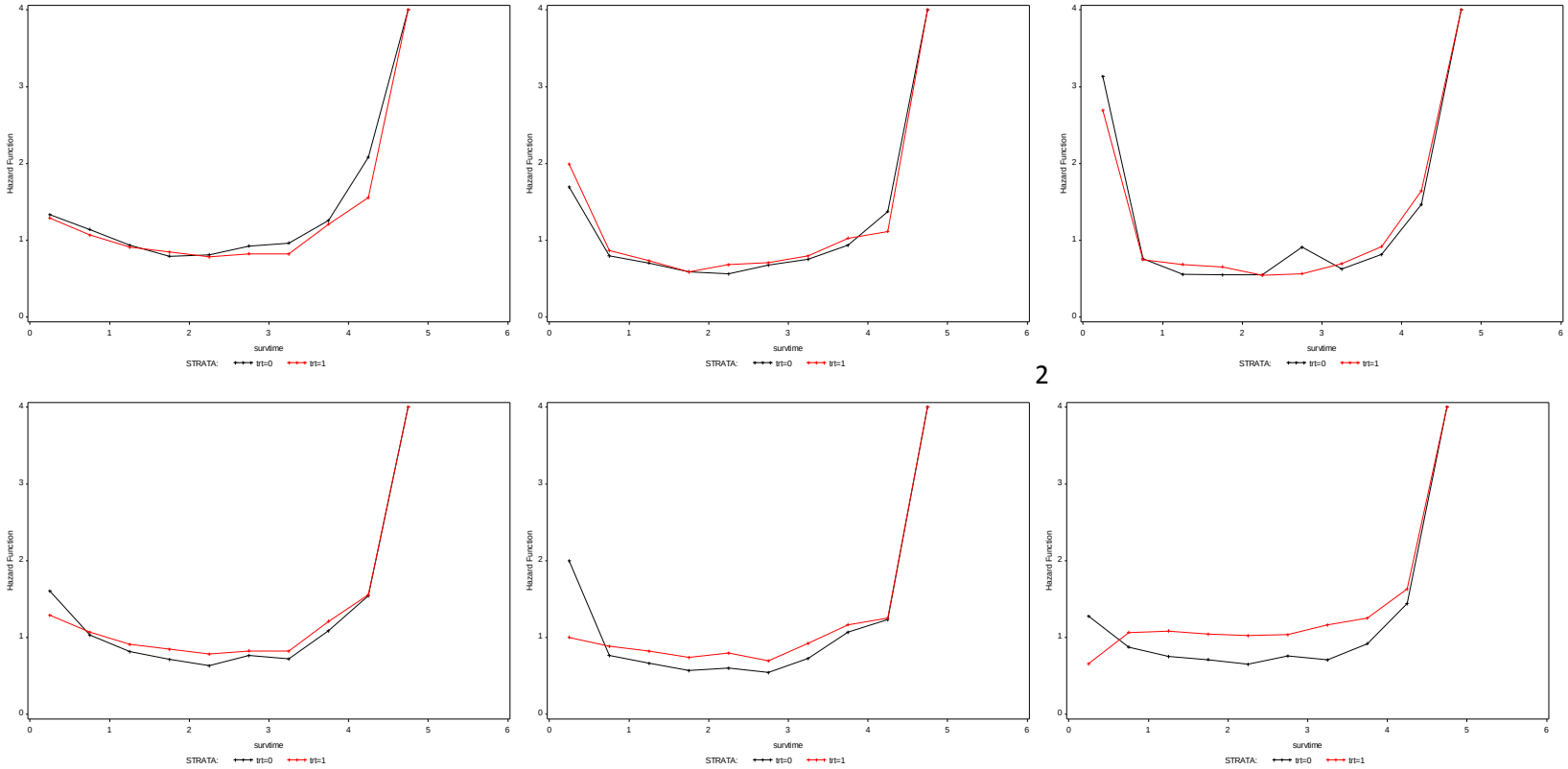


Figure E. Six Patterns of Hazard Functions by Time. The control groups are shown in red while investigational groups are shown in blue. Top Left Panel (Constant – Null); Top Middle (Constant – Alternative); Top Right (Decreasing); Bottom Left (Increasing); Bottom Middle (Diverging); Bottom Right (Crossing).

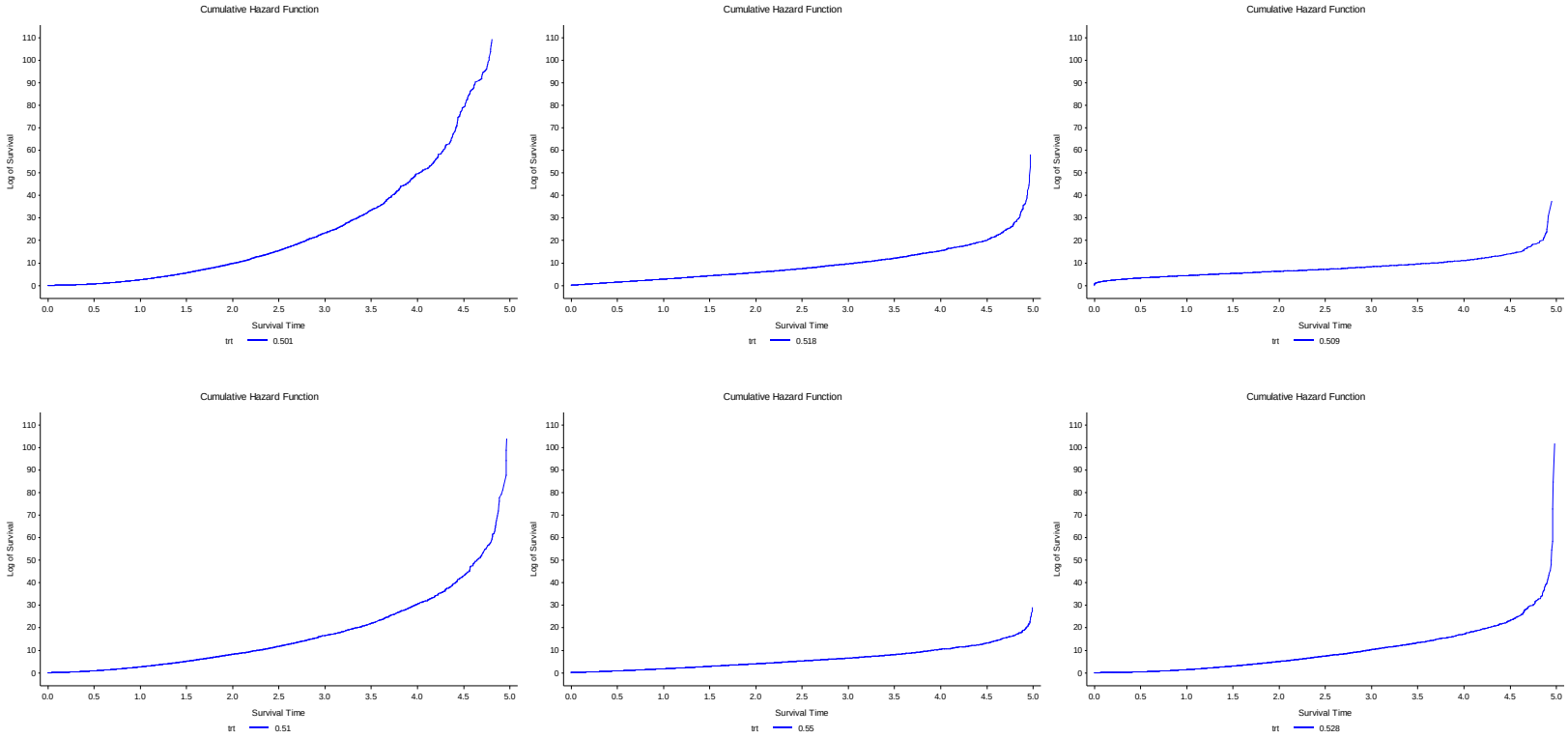


Figure F. Six Patterns of the Cumulative Hazard Functions by Time. The control groups are shown in red while investigational groups are shown in blue. Top Left Panel (Constant – Null); Top Middle (Constant – Alternative); Top Right (Decreasing); Bottom Left (Increasing); Bottom Middle (Diverging); Bottom Right (Crossing).

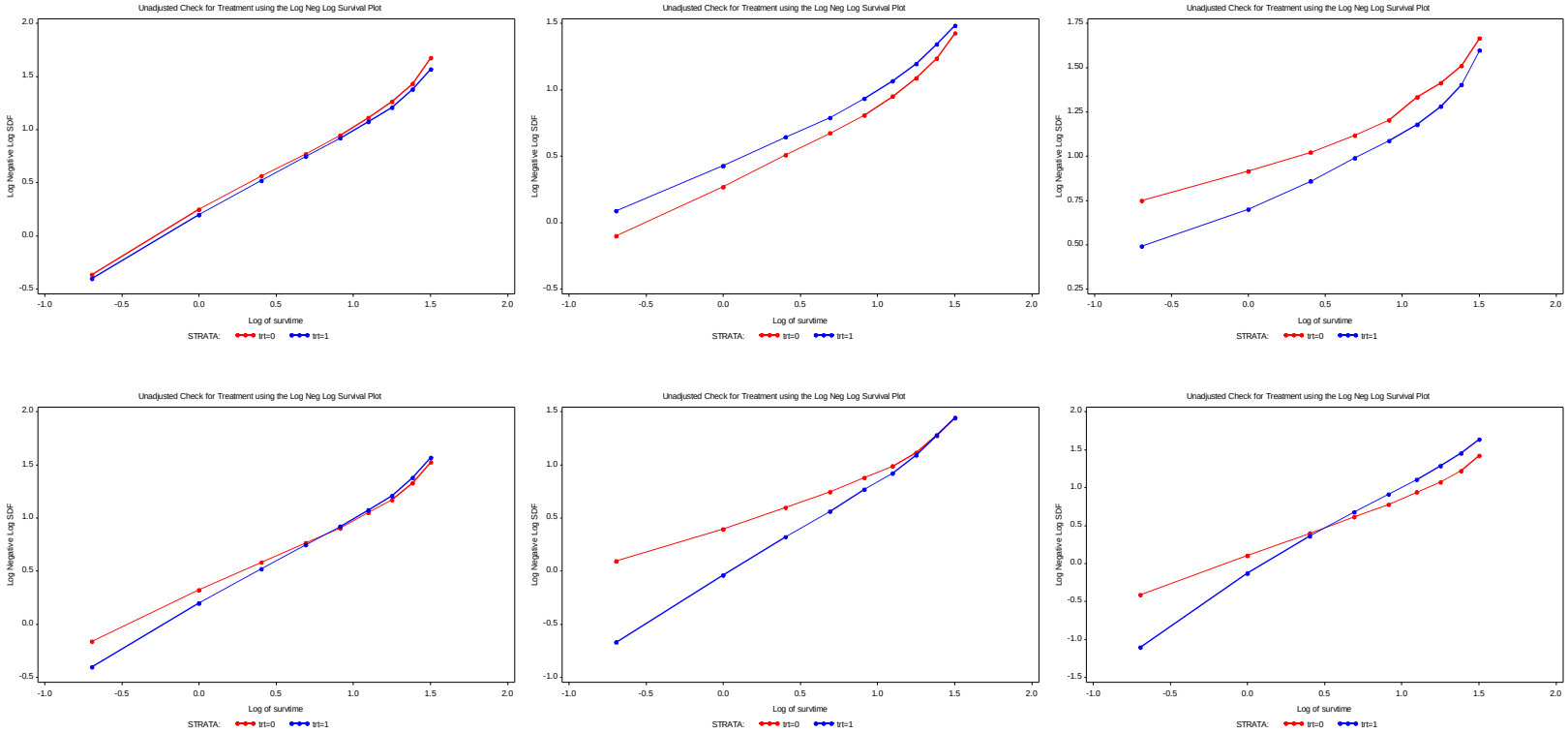


Figure G. Six Patterns of the Unadjusted Log Cumulative Hazard Functions by Log Time. The control groups are shown in red while investigational groups are shown in blue. Top Left Panel (Constant – Null); Top Middle (Constant – Alternative); Top Right (Decreasing); Bottom Left (Increasing); Bottom Middle (Diverging); Bottom Right (Crossing).

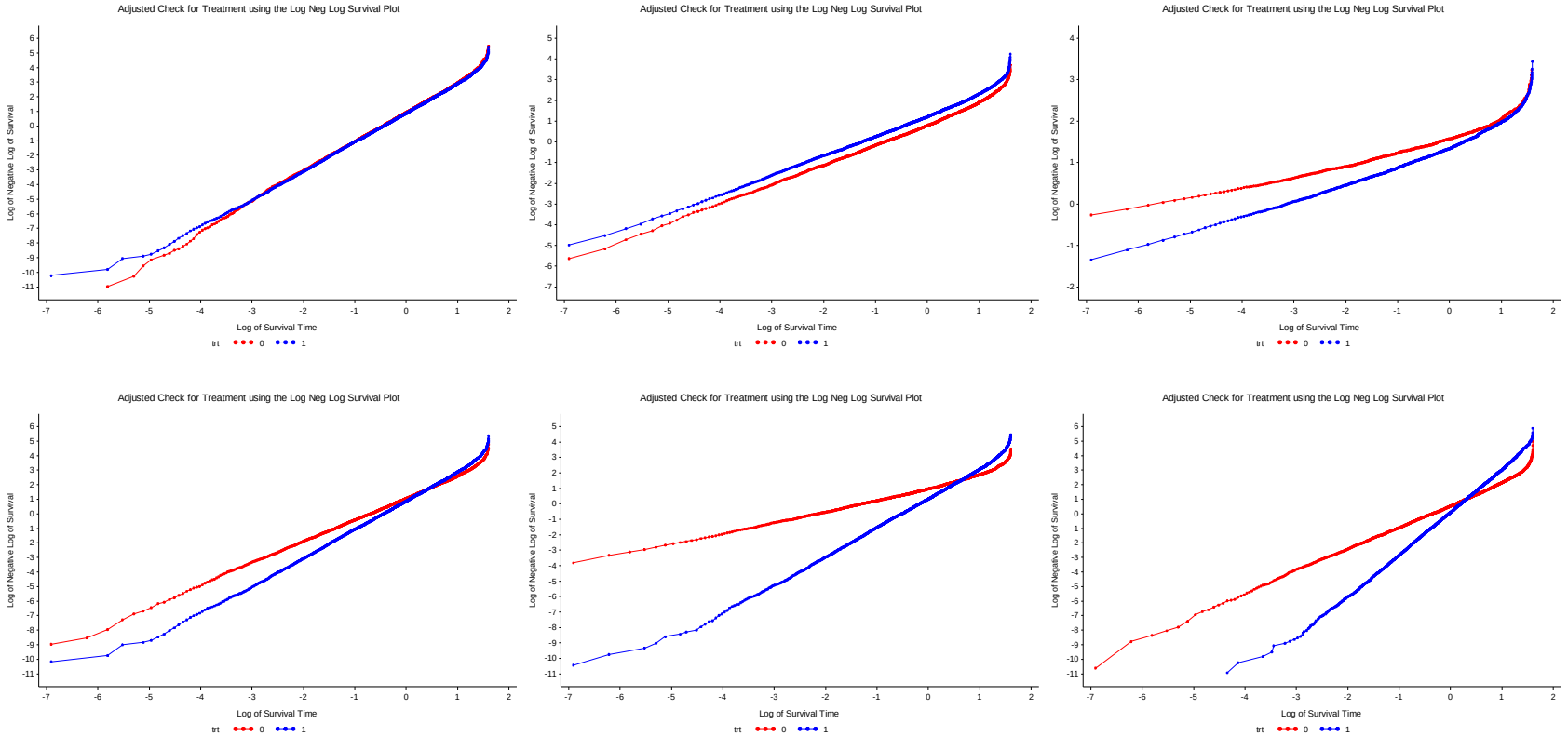


Figure H. Six Patterns of the Adjusted Log Cumulative Hazard Functions by Log Time. The control groups are shown in red while investigational groups are shown in blue. Top Left Panel (Constant – Null); Top Middle (Constant – Alternative); Top Right (Decreasing); Bottom Left (Increasing); Bottom Middle (Diverging); Bottom Right (Crossing).

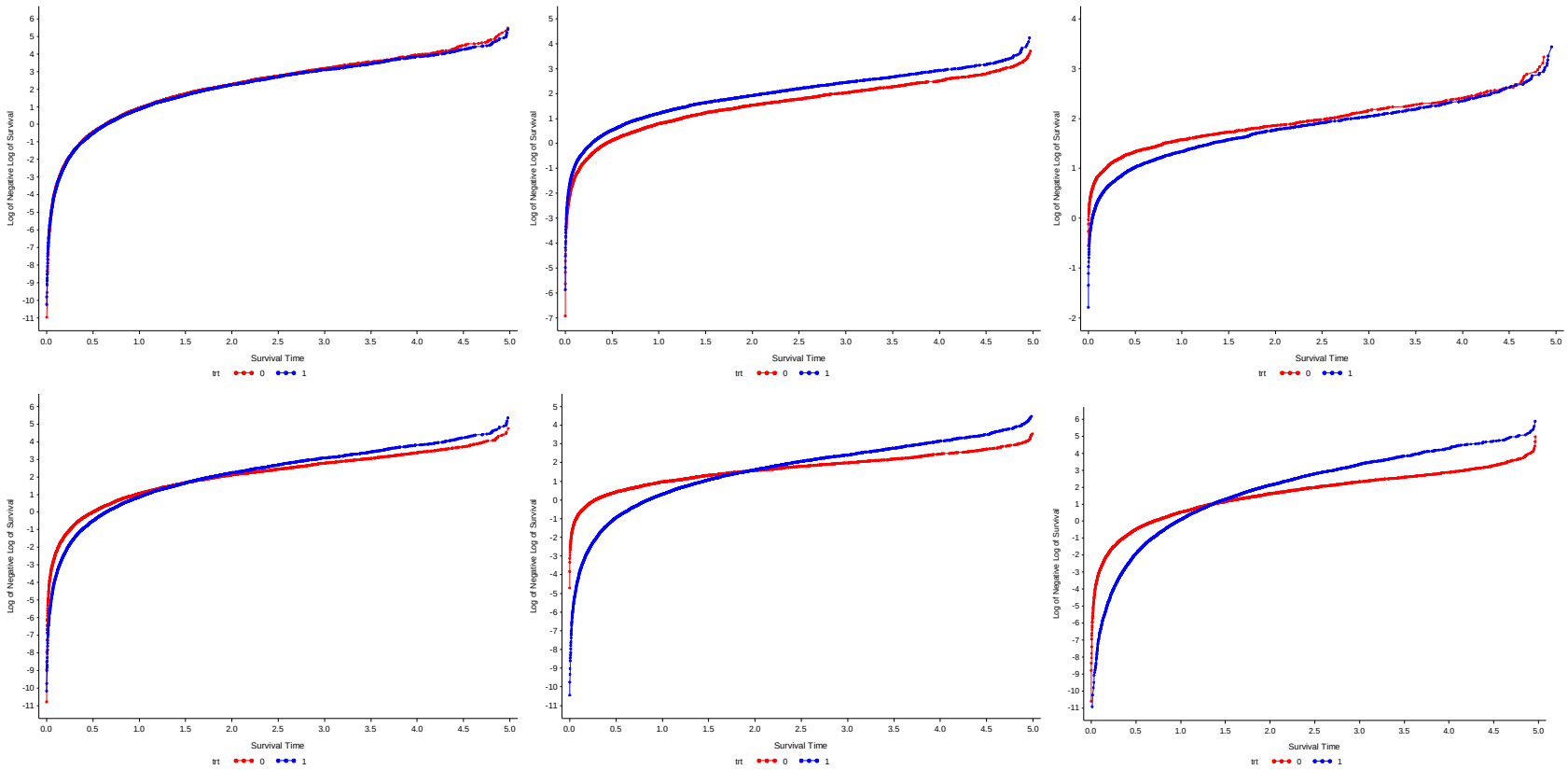


Figure I. Six Patterns of the Adjusted Log Cumulative Hazard Functions by Time. The control groups are shown in red while investigational groups are shown in blue. Top Left Panel (Constant – Null); Top Middle (Constant – Alternative); Top Right (Decreasing); Bottom Left (Increasing); Bottom Middle (Diverging); Bottom Right (Crossing).

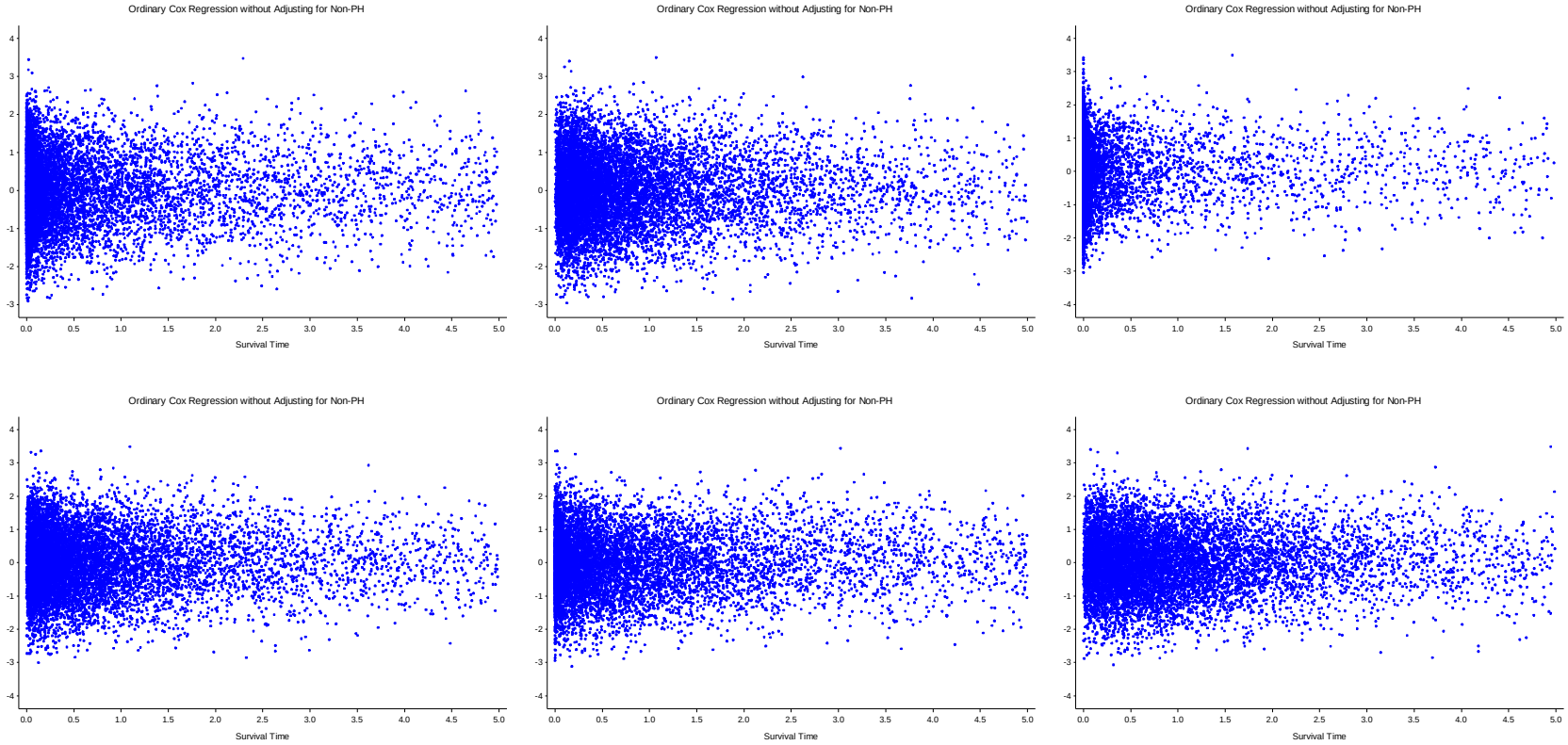


Figure J. Schoenfeld Residuals for a Continuous Variable known to follow the Proportional Hazards Assumption by Survival Time. The control groups are shown in red while investigational groups are shown in blue. Top Left Panel (Constant – Null); Top Middle (Constant – Alternative); Top Right (Decreasing); Bottom Left (Increasing); Bottom Middle (Diverging); Bottom Right (Crossing).

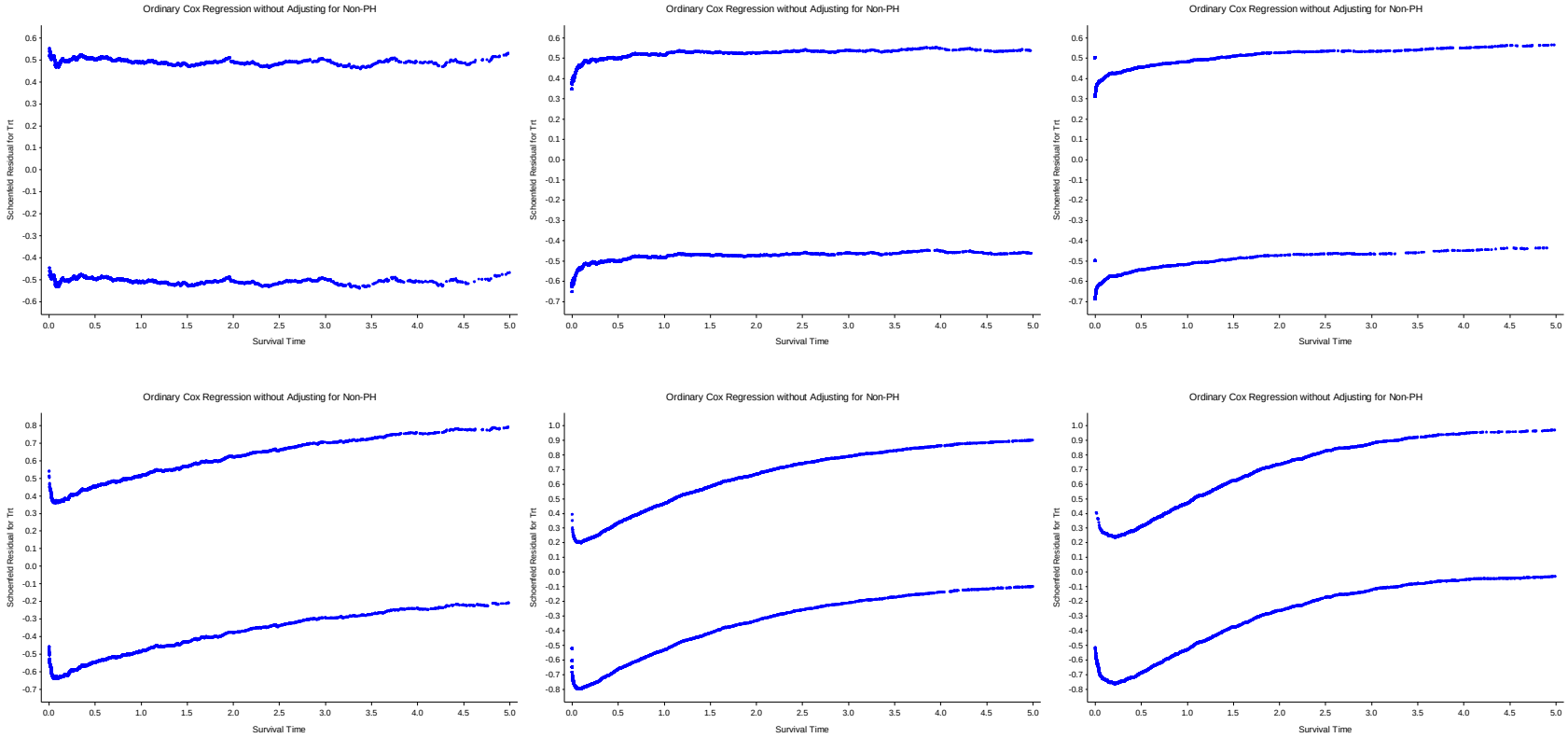


Figure K. Schoenfeld Residuals for a Categorical Variable under different Proportional Hazards Assumptions by Survival Time. The control groups are shown in red while investigational groups are shown in blue. Top Left Panel (Constant – Null); Top Middle (Constant – Alternative); Top Right (Decreasing); Bottom Left (Increasing); Bottom Middle (Diverging); Bottom Right (Crossing)

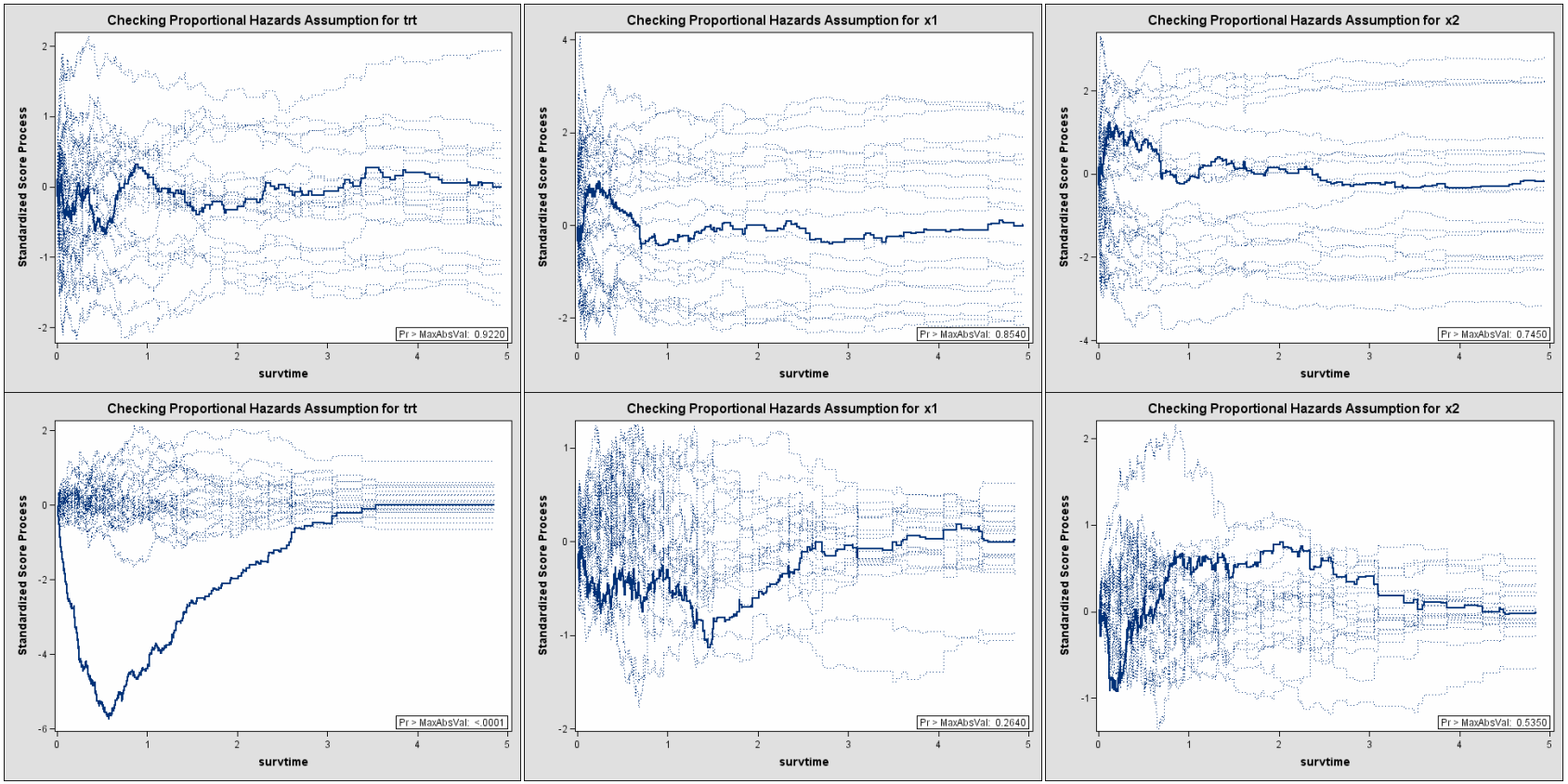


Figure L. Standardized Score Processes for three covariates by Survival Time. The top panels represent the 3 covariate processes from the Alternative case. The bottom panels are the processes from the Crossing case.