Paper AS20

Practical Survival Analysis Techniques for Clinical Trials Using SAS

Xue Ding, Bristol Myers Squibb, Seattle USA

Tianchen Xu, Bristol Myers Squibb, Madison USA

Zihan Lin, Bristol Myers Squibb, Madison USA

ABSTRACT

In clinical trials with time-to-event endpoints, survival analysis is commonly employed to evaluate outcomes such as progression-free survival (PFS) or overall survival (OS), utilizing different methods under proportional hazards or non-proportional hazards scenarios. This tutorial presents a comprehensive, hands-on guide for biostatisticians and statistical programmers applying survival analysis methods to clinical trial data using SAS. We demonstrate primary analyses using the stratified log-rank test and Cox proportional hazards model, including diagnostic tests for the proportional hazards assumption. For scenarios with non-proportional hazards, we highlight advanced methods such as the Max-Combo test and restricted mean survival time (RMST). The tutorial also covers methodological approaches for estimating treatment effects in studies affected by crossover, including the Rank Preserving Structural Failure Time (RPSFT) model, two-stage Accelerated Failure Time (2SAFT) model, and Inverse Probability of Censoring Weights (IPCW). Detailed SAS macros for implementing each method are available from the corresponding author upon request.

INTRODUCTION

This tutorial is designed to demonstrate the application of survival analysis methods to sample datasets using SAS, providing a hands-on guide for biostatisticians and statistical programmers. The primary analysis methods include the stratified log-rank test and Cox proportional hazards model, along with diagnostic testing for the proportional hazards (PH) assumption to ensure appropriate model application. For scenarios involving non-proportional hazards, advanced techniques such as the Max-Combo and restricted mean survival time (RMST) are emphasized to address the complexities of time-varying effects. Additionally, for studies impacted by treatment crossover, this tutorial also introduces robust methodologies such as the Rank Preserving Structural Failure Time (RPSFT) model, two-stage Accelerated Failure Time (2SAFT) model, and Inverse Probability of Censoring Weights (IPCW) approach. By integrating these methodologies with practical SAS coding applications, this tutorial offers a comprehensive guide to addressing diverse survival analysis challenges.

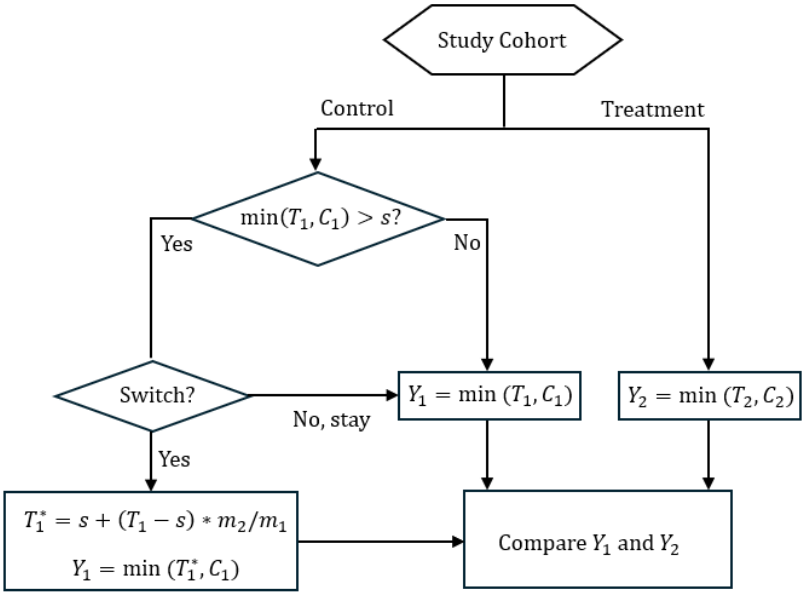
DATA GENERATION

Sample datasets are constructed based on *PowerSwitchingTrial* in Deng 2023, which is freely available at https://github.com/darwin-hub/PowerSwitchingTrial.

The survival time in control and treatment groups, denoted by T_1 and T_2, follow exponential distributions with medians of m_1 and m_2. The censoring time in two groups, denoted by C_1 and C_2, is defined as the time from randomization to dropout or the end of study. For ethical considerations, patients in the control group can switch to the treatment group if the treatment effect is not desired or

patients develop progression. The switching time, denoted by s, is the time from randomization to the time when a patient switches with a probability p. The flowchart of *PowerSwitchingTrial* in Figure 1 illustrates the construction of observed survival time Y_1 and Y_2.

Figure 1: Flowchart of Survival Data with Switching



Two datasets were generated for this guidance:

- *Surv_1* dataset: subject-level data of 500 subjects with age, gender, treatment arm, overall survival time, censoring indication and switching time.
- *Surv_2* dataset: structured in the counting process style of input based on *Surv_1* dataset, where an observation corresponds to a time interval until survival time is reached.

Description of variables:

- Subject: unique subject ID.
- Age: age of subject.
- Gender: gender of subject.
- Arm: specifying which treatment the subject received.
- Time: overall survival time in months.
- Status: censoring indicator for overall survival time. Equal to 1 if subject died and 0 otherwise.
- Switch: cross over time in number of months.
- T_start (only in *Surv_2*): beginning of the time interval in number of months.
- T_end (only in *Surv_2*): end of the time interval in number of months.
- Hgb (only in *Surv_2*): blood hemoglobin in g/dL (time varying covariates).

Variables in Creation Order						
#	Variable	Type	Len	Format	Informat	Label
1	Subject	Num	8	BEST.		Subject ID
2	Age	Num	8	BEST.		Age
3	Gender	Char	6	$6.	$6.	Gender
4	Arm	Char	9	$9.	$9.	Treatment Arm
5	Time	Num	8	BEST.		Overall Survival
6	Status	Num	8	BEST.		Censoring Indicator
7	Switch	Num	8	BEST.		Crossover Time
8	T_start	Num	8	BEST.		Beginning of Time Interval
9	T_end	Num	8	BEST.		End of Time Interval
10	Hgb	Num	8	BEST.		Hemoglobin

PRIMARY ANALYSIS

Stratified log-rank test

The stratified log-rank test is an extension of the standard log-rank test used to compare survival distributions with adjusting for potential confounding by one or more stratification variables. The test assumes that the hazard ratio between groups is constant across strata but allows the baseline hazard to vary by stratum. Suppose there are strata indexed by $k = 1, 2, \dots, K$. The overall stratified log-rank test statistics combine the information across all strata:

$$Z = \frac{\sum_{k=1}^{K}(O_{1k} - E_{1k})}{\sqrt{\sum_{k=1}^{K} V_{1k}}}$$

Where O_{1k} is the observed number of events in group 1, E_{1k} is the expected number of events under the null hypothesis of no treatment difference, and V_{1k} is the variance of $O_{1k} - E_{1k}$. Under the null hypothesis, Z approximately follows a standard normal distribution.

The SAS code uses PROC LIFETEST procedure to conduct survival analysis with log-rank test, the STRATA statement specifies gender as the stratification variable to allow the analysis to adjust for gender differences across groups. The GROUP option specifies arm as the grouping variable.

```
proc lifetest data=surv_1;
time Time*Status(0);
strata gender / group = arm;
run;
```

Stratified Test of Equality over Group			
Test	Chi-Square	DF	Pr > Chi-Square
Log-Rank	12.6519	1	0.0004
Wilcoxon	11.0244	1	0.0009

The Chi-Square and P-value associated with Log-Rank test are used to determine if there are significant differences in survival between treatment and control group.

Stratified Cox proportional hazards model

The stratified Cox proportional hazards model allows for different baseline hazard functions across strata while assuming a common effect of covariates across those strata. Suppose the data are divided into K strata indexed by $k = 1, 2, \dots, K$. The hazard function for individual i in stratum k at time t is:

$$\lambda_i^{(k)}(t) = \lambda_0^{(k)}(t)\exp\left(\beta^T X_i\right)$$

where $\lambda_0^{(k)}(t)$ is the baseline function to stratum k, β is the vector of coefficients, and X_i is the vector of covariates. The partial likelihood for estimating β is the product of the partial likelihood across all strata:

$$L(\beta) = \prod_{k=1}^{K} \prod_{i \in D_k} \frac{\exp\left(\beta^T X_i\right)}{\sum_{j \in R_{k(t_i)}} \exp\left(\beta^T X_j\right)}$$

Where D_k is the set of individuals who experienced an event in stratum k, and $R_{k(t_i)}$ is the risk set in stratum k at time t_i.

The SAS code uses PROC PHREG procedure to fit Cox proportional hazards model. The CLASS statement defines categorical variables and the STRATA statement incorporates gender as a stratification variable. Additionally, the HAZARDRATIO statement computes the hazard ratio between treatment and control group, with confidence intervals using Wald method.

```
proc phreg data=surv_1;
class gender arm(ref='control');
strata gender;
model Time*Status(0)=arm/risklimits;
hazardratio arm / cl = wald;
run;
```

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	12.5840	1	0.0004
Score	12.6519	1	0.0004
Wald	12.4732	1	0.0004

Analysis of Maximum Likelihood Estimates										
Parameter		DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	95% Hazard Ratio Confidence Limits		Label
Arm	treatment	1	-0.41539	0.11762	12.4732	0.0004	0.660	0.524	0.831	Treatment Arm treatment

The Chi-Square and P-value of Likelihood Ratio, Score and Wald test can be used to determine if there is a significant difference in survival between treatment and control groups. In addition, the estimated hazard ratio was reported with a 95% confidence interval to measure the difference between two groups.

GRAPHICAL ANALYSIS FOR PH ASSUMPTION

Graphical analysis provides an intuitive method for evaluating the proportional hazards assumption in survival analysis. Two commonly used approaches are log-minus-log plots and Schoenfeld residuals plots, both offer visual insights into potential deviations from the assumption.

For the log-minus-log plot applied to two-sample survival data, the proportional hazards assumption is met when the following relationship holds true:

$$\log(-\log\left(S(t)\right) = \log\left(-\log\big(S_0(t)\big)\right) + \beta x$$

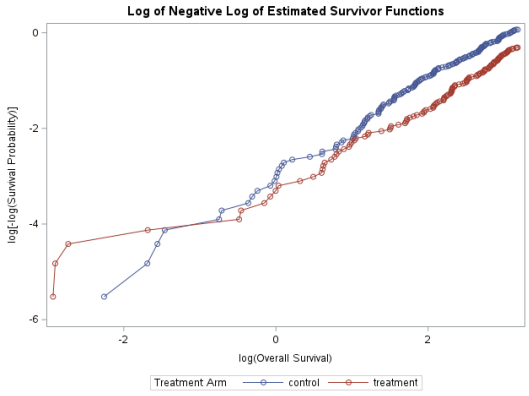
Where $S(t)$ and $S_0(t)$ are survival functions for control and treatment arms, β is the coefficient and x is the indication variable for treatment arm. Parallel curves of two arms indicate that the PH assumption is not violated.

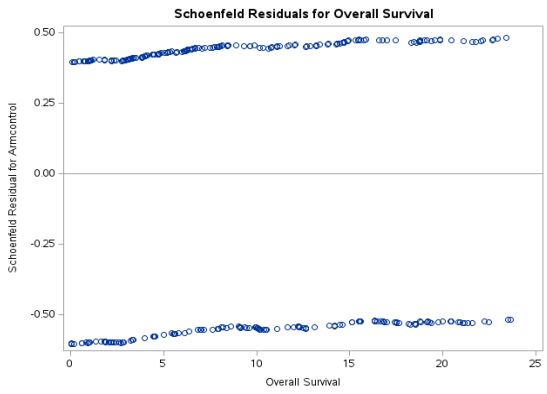
```
proc lifetest data=surv_1 plots=(lls);
time Time*Status(0);
strata arm;
run;
```

Schoenfeld residuals are defined as $x_i - E[X|t_i]$ where x_i is the value of covariate for individual who experienced the event at time t_i, $E[X|t_i]$ is the expected value of that covariate over the risk set at time t_i, weighted by partial likelihood. By plotting Schoenfeld residuals against time, patterns or trends can reveal violations of the assumption, highlighting any time-dependent effects in the covariates.

```
proc phreg data=surv_1;
class arm;
model Time*Status(0) = arm;
output out=out ressch=schres;
run;

proc sgplot data=out;
scatter y=schres x=Time / name='scatter';
title "Schoenfeld Residuals for Overall Survival";
refine 0;
run;
```





NON-PROPORTIONAL HAZARDS ANALYSIS

Restricted Mean Survival Time (RMST)

RMST (Royston, 2013) is defined as the expected survival time up to a pre-specified time point τ, and it provides an alternative to the hazard ratio by directly quantifying the average survival benefit within a fixed time window. It tests the areas under the survival curve from time 0 to τ:

$$RMST(\tau) = \int_0^{\tau} S(t)dt$$

where $S(t)$ is the survival function, which can be estimated using the Kaplan-Meier estimator $\hat{S}(t)$. The difference in RMST is a common measure of absolute treatment effect:

$$\theta = RMST_1(\tau) - RMST_2(\tau)$$

The SAS code utilizes PROC LIFETEST procedure to calculate RMST. The DIFF=ALL option computes differences in RMST between each pair of treatment arms. The default RMST method uses min-max value for τ unless it is pre-specified.

```
proc lifetest data=surv_1 rmst;
time Time*Status(0);
strata arm / diff = all;
run;
```

Restricted Mean Survival Time Comparisons

Stratum Comparison		Difference	Standard Error	Chi-Square	Pr > ChiSq
control	treatment	-2.93567	0.7634	14.7891	0.0001

The Chi-Square and P-value are reported to determine the significant difference between two RMSTs. And the absolute difference is also reported.

Max-Combo Test

The Max-combo test (Karrison, 2016) combines several weighted log-rank tests. For patients $i = 1, 2, \dots, I$, the weighted log-rank test is defined as:

$$Z_w = \frac{\sum_{i=1}^{I} w(t_i)(O_{1i} - E_{1i})}{\sqrt{\sum_{i=1}^{I} V_{1i}}}$$

Common choices for weight functions include the Fleming-Harrington family $G^{\rho,\gamma}$ with $w(t) = \hat{S}(t)^{\rho}(1 - \hat{S}(t))^{\gamma}$, where $\hat{S}(t)$ is the Kaplan-Meier estimate. Consider a set of weighted log-rank tests $Z_{0,0}$, $Z_{0,1}$, $Z_{1,0}$, $Z_{1,1}$, corresponding to the weight functions $G^{0,0}, G^{0,1}, G^{1,0}$ and $G^{1,1}$, respectively. The Max-combo test is then defined as:

$$Z_{max} = \max(Z_{0,0}, Z_{0,1}, Z_{1,0}, Z_{1,1})$$

The Max-combo test provides robustness across various non-proportional hazard patterns.

The SAS macro %combo_wlr implements the calculation of maximum weighted log-rank test. The inputs are survival dataset, grouping variable, survival time, censoring variable and a sequence of weight vector.

```
%combo_wlr(data=surv_1, group=arm, time=Time, event=Status, weights=%str(0,0 1,0 0,1 1,1));
```

Weighted log-rank tests

Test	Z statistic	P
Fleming(0,0)	3.54507	0.0004
Fleming(1,0)	3.71921	0.0002
Fleming(0,1)	2.47705	0.0132
Fleming(1,1)	2.97755	0.0029

Combination test

Z max	P
3.71921	0.0004

The test statistics and p-values for each individual FH test are reported. The Z max and p-value for the combination test are used to determine the significance in survival between two groups. And the p-value from the combination test was adjusted for multiplicity.

TREATMENT CROSSOVER

Rank Preserving Structure Failure Time (RPSFT)

The RPSFT model (Ishak, 2014) assumes that there exists a common factor for patients who switch from control to treatment group which preserves the ranking of failure time. The counterfactual survival time T without treatment for patients in control group can be transformed into the observed failure time by the model:

$$T = T_{off} + e^{\psi} T_{on}$$

where T_{off} and T_{on} indicate the time 'off' and 'on' treatment group. e^{ψ} is the acceleration factor.

The estimation of φ can be implemented with various methods, including log-rank test, Weibull test and Cox model, the optimal value is the value that set test statistics equal to 0, which implies treatment group not related to survival time. Survival time for patients who switch from control to treatment group can be adjusted using the optimal value of φ. Cox model fitting and hazard ratio estimation are performed based on these adjusted survival times.

Informative censoring poses challenges for the RPSFT model in accurately estimating counterfactual survival times. Re-censoring is implemented to ensure that the counterfactual survival time is censored at the same time point as the original observed data, thereby preventing it from extending beyond the original censoring time.

The SAS macro %rpsft implements the calculation of the optimal φ and cox model fitting with adjusted survival time. The inputs are survival dataset, survival time, censoring variable, treatment group, switching time, lower bound, upper bound and interval of φ and re-censoring.

```
%rpsft(data=surv_1, aval=time, cnsr=status, trt=arm, xo_time=switch, psi_start=-1, psi_end=1, psi_step=0.1, recnsr='N');
```

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	15.0743	1	0.0001
Score	15.1955	1	<.0001
Wald	14.9401	1	0.0001

Analysis of Maximum Likelihood Estimates									
Parameter		DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	95% Hazard Ratio Confidence Limits	Label
Arm	treatment	1	-0.45517	0.11776	14.9401	0.0001	0.634	0.504 0.799	Treatment Arm treatment

```
%rpsft(data=surv_1, aval=time, cnsr=status, trt=arm, xo_time=switch, psi_start=-1, psi_end=1, psi_step=0.1, recnsr='Y');
```

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	16.8964	1	<.0001
Score	16.9124	1	<.0001
Wald	16.4731	1	<.0001

Analysis of Maximum Likelihood Estimates										
Parameter		DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	95% Hazard Ratio Confidence Limits		Label
Arm	treatment	1	-0.56486	0.13917	16.4731	<.0001	0.568	0.433	0.747	Treatment Arm treatment

The output presents results from the Cox model and hazard ratio estimation.

Two-stage Accelerated Failure Time (2SAFT)

Similarly, the two-stage AFT model (Latimer, 2014 and Latimer, 2016) assumes the counterfactual survival time by:

$$T = T_A + e^{\psi} T_B$$

where T_A is the time from randomization to treatment crossover, T_B is the time from crossover to death or censoring. In the first stage, an accelerated failure model (e.g. Lognormal, Weibull or Gamma distribution) is fitted to estimate φ.

The second stage is to recalculate the counterfactual survival time for patients who switch from control to treatment group using the estimated φ. Cox model fitting and hazard ratio estimation are performed based on these adjusted counterfactual survival times.

As with the RPSFT model, re-censoring has been incorporated into the 2SAFT method and is available as an optional feature.

The SAS macro %twosaft implements the estimation of φ and cox model fitting with adjusted counterfactual survival times. The inputs are survival dataset, survival time, censoring variable, treatment group, switching time, accelerated failure model and re-censoring.

```
%twosaft(data=surv_1, aval=time, cnsr=status, trt=arm, xo_time=switch, af_dist=weibull, recnsr='N');
```

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	16.2705	1	<.0001
Score	16.4230	1	<.0001
Wald	16.1258	1	<.0001

Analysis of Maximum Likelihood Estimates										
Parameter		DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	95% Hazard Ratio Confidence Limits		Label
Arm	treatment	1	-0.47310	0.11781	16.1258	<.0001	0.623	0.495	0.785	Treatment Arm treatment

```
%twosaft(data=surv_1, aval=time, cnsr=status, trt=arm, xo_time=switch, af_dist=weibull, recnsr='Y');
```

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	16.8186	1	<.0001
Score	16.7926	1	<.0001
Wald	16.2836	1	<.0001

Analysis of Maximum Likelihood Estimates										
Parameter		DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	95% Hazard Ratio Confidence Limits		Label
Arm	treatment	1	-0.61077	0.15136	16.2836	<.0001	0.543	0.404	0.730	Treatment Arm treatment

Inverse Probability of Censoring Weighting (IPCW)

The IPCW method applies weights to construct a pseudo-population for the control arm, representing a counterfactual scenario in which no crossover occurred. This model relies on assumption that all confounders are measured and that there are no unmeasured confounders. To estimate the time-dependent weights, both baseline covariates and time-varying covariates are incorporated in the weight calculation.

Assuming time-varying covariate was involved, the weight function for a patient i at time point t_j is given by:

$$w_{ij} = \prod_{k=1}^{j} \frac{P(C_i > t_k | C_i > t_{k-1}, X_i)}{P(C_i > t_k | C_i > t_{k-1}, X_i, Z_i)}$$

Where C_i is censoring time, X_i is baseline covariates and Z_i is time-varying covariates.

A Cox model can be used to estimate the probability of remaining uncensored for both numerator and denominator of the weights. For illustration purposes, we consider only the situation in which patients are only allowed to switch from the control arm to the treatment arm, thus weights are estimated only for participants in the control arm, whereas those in the treatment arm are assigned a constant weight of 1. Two models will be run for weight estimation, the estimation of numerator is based solely on baseline covariates, whereas the estimation of denominator incorporates both baseline and time-varying covariates, and the estimated weight is calculated as the numerator divided by the denominator. After the weights are calculated, the survival time and corresponding weights can be incorporated into a Cox model to estimate the hazard ratio. Stabilized weights may also be used to reduce the influence of extreme values, and a simple implementation approach is to truncate the estimated weights at specified quantiles.

The SAS macro %ipcw implements the estimation of weights and Cox model fitting. The inputs are survival dataset, survival time, treatment group, censoring variable, switching time, baseline covariates, start and end of time interval, and time-varying covariates.

```
%ipcw(data=surv_2, subject=subject, trt=arm, censor=status, xo_time=switch, bl_covar=gender, t_start=t_start, t_end=t_end, tv_covar=hgb);
```

Analysis of Maximum Likelihood Estimates
with Model-Based Variance Estimate

Parameter		DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	95% Hazard Ratio Confidence Limits		Label
Arm	control	1	0.29448	0.07940	13.7561	0.0002	1.342	1.149	1.568	Treatment Arm control

Analysis of Maximum Likelihood Estimates
with Sandwich Variance Estimate

Parameter		DF	Parameter Estimate	Standard Error	StdErr Ratio	Chi-Square	Pr > ChiSq	Hazard Ratio	95% Hazard Ratio Confidence Limits		Label
Arm	control	1	0.29448	0.11461	1.443	6.6022	0.0102	1.342	1.072	1.681	Treatment Arm control

CONCLUSION

This tutorial presents a comprehensive framework for applying survival analysis techniques to clinical trial data using SAS. Standard methods such as the stratified log-rank test and Cox proportional hazards

model are demonstrated, along with diagnostic tools for assessing the proportional hazards assumption. For scenarios involving non-proportional hazards, alternative approaches including restricted mean survival time (RMST) and the Max-Combo test are introduced to address time-varying treatment effects.

To account for treatment crossover, the tutorial outlines three adjustment methodologies—Rank Preserving Structural Failure Time (RPSFT), two-stage Accelerated Failure Time (2SAFT), and Inverse Probability of Censoring Weighting (IPCW)—each supported by detailed SAS macro implementations. These methods enhance the accuracy of treatment effect estimation in trials where crossover is ethically or clinically necessary.

By integrating methodologies with practical coding, this resource supports the implementation of robust survival analyses in clinical research settings.

REFERENCE

Deng, L., Hsu, CY. & Shyr, Y. Power and sample sizes estimation in clinical trials with treatment switching in intention-to-treat analysis: a simulation study. BMC Med Res Methodol 23, 49 (2023).

Grambsch PM, Therneau TM. Proportional hazards tests and diagnostics based on weighted residuals. Biometrika. 1994 Aug;81(3):515-526.

Ishak KJ, Proskorovsky I, Korytowsky B, Sandin R, Faivre S, Valle J. Methods for adjusting for bias due to cross over in oncology trials. Pharmacoeconomics. 2014 Jun 01;32(6):533-546.

Karrison TG et al. Versatile tests for comparing survival curves based on weighted log-rank statistics. Stata J. 2016; 16(3):678-690.

Latimer NR, Abrams KR, Lambert PC, Crowther MJ, Wailoo AJ, Morden JP, Akehurst RL and Campbell MJ. Adjusting for treatment switching in randomized controlled trials - A simulation study and a simplified two-stage method. Statistical Methods in Medical Research. 2014 Nov21;26(2):724-751.

Latimer NR, Abrams KR, Lambert PC, Morden JP, Crowther MJ. Assessing methods for dealing with treatment switching in clinical trials: A follow-up simulation study. Statistical Methods in Medical Research. 2016 Apr 25.

Royston, P., Parmar, M.K. Restricted mean survival time: an alternative to the hazard ratio for the design and analysis of randomized trials with a time-to-event outcome. BMC Med Res Methodol 13, 152 (2013).

CONTACT INFORMATION

Your comments and questions are valued and enrouraged. Contact the author at:

Author Name: Xue Ding

Company: Bristol Myers Squibb

Email: xue.ding@bms.com