International Journal of Progressive Sciences and Technologies (IJPSAT)
ISSN: 2509-0119.

http://ijpsat.ijsht-journals.org Vol. 10 No. 2 September 2018, pp. 284-293

A Simulation Study to Evaluate the Performance of Extended Cox model in Testing Treatment Effect with Possible Non-proportional Hazards

Belay Belete Anjullo

Department of Statistics, Arba Minch University,
P.O. Box: 21, Arba Minch, Ethiopia

Abstract:
Background: Many randomized clinical studies with long-term follow-up regularly measure time-to-event outcomes, such as survival time to compare an experimental treatment with a standard treatment or placebo control. In this comparison, one tests whether the two treatments have the same survival function or equivalently the same hazard function over a given time period in order to evaluate effect of treatment. However, when comparing treatments in terms of their time to event distribution, there may be reason to believe that the hazard curves will cross, and in such cases standard comparison techniques could lead to misleading results. It was shown that extended Cox model performed well to test effect of treatment in case of crossing survival curves at 20% and 40% administratively censored time. However, most of the studies did not assess the statistical properties of this approach under high censoring rates.
Objective: In this paper, the main objective of the study was to evaluate whether the power and type I error rate of extended Cox model was robust with variations in censoring rates under five possible treatment effects based on data from simulations.
Methodology: a simulation study was designed to compare the performance of the extended Cox and the data are replicated 1000 times to estimate type I error rate and empirical power of the test.
Results: type I error rates for the test from the model approached the nominal level of 0.05 at 10% and 30% of censoring rates regardless of sample size per treatment group considered in the study, however, type I error rates are slightly inflated at 70% of censoring rate. Moreover, with respect to powers of the test, extended Cox model performed reasonably well with powers of test above 80% in case of crossing survival curves under censoring rates of 10% and 30% to test treatment effect without including baseline covariates in the model. Although 70% censoring rate appeared to have influence on the power of test, test from extended Cox model exhibits moderate power in this situation, particularly for sample of size 200 per treatment group. Furthermore, extended Cox model with heavy side function performed well under the early treatment effect in which hazards is expected to crosses. Using the extended Cox model with baseline covariate in case of crossing survival curves did not generally yield dramatic decrease in power.
Conclusions: 70% censoring rate appeared to have influence on the power of test and type I error rate although test exhibits moderate power for sample of size 200 per treatment group.

Keywords - Simulation, Type I Error Rate, Cox Proportional Hazards, Power of The Test, Extended Cox Model

I. INTRODUCTION

1.1. Background

Survival analysis has become one of the most widely used statistical tools for analyzing clinical research data. It is specifically concerned with time to event data and is of particular value because of its intrinsic ability to handle censored observations. In many randomized clinical trials includes right censored time to event data, comparing an experimental treatment with a standard treatment or placebo control in order to evaluate treatment e□ect [1],[2]. In this case, one tests whether the two treatments have the same survival function or equivalently the same hazard function over a given follow up time[3]. The log-rank test is commonly used test statistic for the comparison. However, in these trials the characteristics of the patient and of the tumours that are known before treatment are also recorded.

Corresponding Author: Belay Belete Anjullo

Therefore, Cox proportional hazards model is the most common choice to compare the effect of treatment with advantages of adjusting for baseline and prognostic covariate(s)[4].

However, Cox model has the assumption of proportional hazards, meaning that the ratio of the hazards for treatment versus control is constant over time [5]. Although not as implicitly assumed as in the Cox regression model, the validity of the log-rank test is also sensitive to the assumption that the hazard ratios for treatment versus control do not change appreciably over time [6]. When studying survival data over a short period of time, the proportional hazards assumption is often a reasonable one. However, in cancer clinical trials with long-term follow-up, it often happens that the hazard ratio changes over time. In the beginning of the study for instance, the experimental treatment may yield better survival, but this effect may be reversed after some time or vice versa [4]. In such a case, the log-rank test for the difference in survival ratios between the treatments will most likely not be significant, because of the contrasting early and late effects of the treatments. When the proportional hazards assumption is not satisfied for the treatment or for one or more of the covariates, the results of a Cox model will be misleading. In addition, interpretation of the hazards ratio resulting from the Cox proportional hazards model is not easy because it is a weighted average hazards ratio over the observed follow-up time [7], [8].

In the literature, to deal with the issues of non-proportional hazards, the Cox regression model with time-dependent treatment effects was proposed [6]. Klein proposed to compare survival curves at one fixed time point [9],[10]. It was proposed to compare restricted mean survival time at a pre-specified fixed time point[7]. Chen and Tsiatis studied methods for comparing covariate-adjusted restricted mean survival times between two treatment groups [11]. It was also proposed for testing treatment effect by combining weighted log-rank tests and using empirical likelihood [2]. Logan proposed to test two sub hypotheses: the hypothesis of equality of Kaplan-Meier survival difference at a pre-specified time point (t_0) and the hypothesis of no difference in the hazards after t_0 [4].

Moreover, a key characteristic that differentiates survival analysis from other areas in statistics is that survival data are usually censored. Censoring has an impact on the shape of survival curve in a situation when a large number of individuals are censored at a single point of time leading to sudden spurious large jumps or large flat section in survival curve [12]. Maximum censoring occurs in cancer clinical studies where the participants are followed through the duration of the study period. In this case, those that are most likely to be censored are those participants that survive the longest, for two reasons. The first reason is that the longer a participant is in a study, the greater the chance of them getting lost. The second reason is that those who live beyond the time period of the study will all be censored, and since everyone began the study at the same time, the participants that live beyond the time period of the study will also be the people who have the longest survival times [13].

It was shown that extended Cox model performed about the same compared to methods based combination of early/late treatment effects obtained from stopped/left truncated Cox model to test effect treatment in case of crossing survival curves at 20% and 40% administratively censored time [14]. Nonetheless, most of the studies did not evaluate the statistical properties of this approach under varieties of censoring rates specifically at high censoring rate. Hence, in this paper, data from simulations was used to investigate whether the power and type I error rate of extended Cox model was robust with variations in censoring rates under various configurations of survival curves for treatment group.

1.2. Statement of the Problem

In many randomized clinical trials in oncology, a new treatment group is compared with an existing standard treatment or control group based on the hazard ratio (HR), estimated by the Cox proportional hazards model [5]. When hazards are not proportional, the Cox may not be powerful; consequently, different approaches have been proposed as alternative to the Cox model in the case of non-proportional hazards. It was reported that a test statistic from extended Cox model has optimal power to detect differences in the hazard rates, when the hazard rates are non-proportional [14]. However, most studies have not evaluated the statistical properties of this approach under high censoring rate. Therefore, this paper has attempted to answer the following scientific question:- How well extended Cox model work as censoring rate changes in the situation where proportional hazards assumption is not satisfied for variety of sample sizes?

1.3. Objective of the Study

The main objective of this paper was to evaluate the strengths and weaknesses of the extended Cox model in order to test treatment effect in randomized clinical trials

with possible non proportional hazards with variation of censoring rates under 50,100,200 subjects per treatment group.

II. Methodology

2.1. Simulation Design

A simulation study was designed to compare the performance of the extended Cox model in order to test treatment effect in randomized clinical trials for a different possible treatment effects and censoring rates under different sample size in terms of its type I error rate and power. In the simulation design, survival times for treatment groups were generated independently for samples of size 50,100,200 subjects per treatment group with 10%, 30% and 70% of administrative censoring (censoring due to termination of study) using true survival functions presented in Figures 1-5. This was done under five different scenarios such as: in scenario 1) survival curves are assumed to be identical (i.e., no treatment effect under the null hypothesis), 2) survival curves are assumed to have proportional hazards, 3) survival curves are assumed identical at the beginning, then separate as time goes on (late treatment effect), 4) the two survival curves are separate at beginning, but identical as time goes on (early treatment effect) leading to crossing hazards, and 5) survival curves are assumed to cross. In all scenarios survival times are simulated conditioning on the binary covariate which was generated from Bernoulli distribution considering the follow up period of five years and independent of the censoring times. For each scenario, the data are replicated 1000 times which is the most common choices [15]. The type I error rate and empirical power of the test is calculated as the proportion of 1000 repeated random samples in which the null hypothesis is rejected at the nominal alpha of 5% with one-sided test statistics under identical survival curves and four different alternative scenarios, respectively with and without including the covariate in the models. This simulation setting was adopted from a simulation design used to evaluate the performance of the new proposed methods in testing treatment effect, whether or not the underlying proportional hazards assumption was met [14], [16-18]. Simulations and analysis were done using R software version of R3.4.0.

2.2. Testing Treatment Effect based on Extended Cox Model

In cancer clinical trials, which require long-term follow up, a new treatment group is compared with an existing standard treatment or control group based on the hazard ratio estimated by the Cox model in which the hazard ratio often changes over time [2]. In such cases, it is not easy to interpret hazard ratio estimated by the Cox model. Furthermore, when hazard ratio cross, the Cox test is known to lose power. In this situation a way of studying the effect of treatment changes over time by adding a time dependent treatment effects in a Cox proportional hazards model [6], [19-20]. The most straightforward way to model a time dependent treatment effect is by adding interaction terms of the treatment group with $f(t)$ as $\beta_G G(t) = \beta_G G f(t)$, where $f(t)$ is the function of time t with its popular choice can be t or $\log(t)$ or heaviside function that take value 1 for all time point greater than or equal to pre-specified time t_0 or zero otherwise. In this study, for the practicality and comparability of results, heaviside function which is conceptually related with stopped Cox and defined on the median of observed events (t_0) was adopted. In the literature it is stated that, if there is no information about crossing point for hazards the recommended choice is the time point where half of the expected number of event are observed [16]. Gillen and Emerson also suggested the use of equally spaced information time with the goal of balancing loss of statistical power against the potential for early stopping in the situation where there is no prior knowledge of a time varying treatment effect [21]. These are considered as motivations for the choice of time point t_0 in this study. The general form of the extended Cox model with time dependent treatment effect can be written as:

$h(t) = h_0(t)\exp(\beta_E G * (1 - \mathrm{f(t)}) + \beta_L G * \mathrm{f(t)} + X'\beta_X)$, where

$f(t) = \begin{cases} 1, t \geq t_0 \\ 0, t < t_0 \end{cases}$ is called heaviside (step) function, G is treatment groups (1 for treated and 0 for control), X are additional baseline covariate(s), β_E , β_L and β_X are parameters to be estimated representing early, late treatment effects and baseline covariate(s) effects, respectively. The parameters of the model were estimated by maximizing logarithm of partial likelihood via Newton-Raphson iterative procedure [22]. Let's denote $Pvalue_{early}$ and $Pvalue_{late}$ as one-sided p-value to test for the early and late treatment effect with hypothesis $H_{oe}: \beta_E = 0$ versus $H_{1e}: \beta_E < 0$ and $H_{ol}: \beta_L = 0$ versus $H_{1l}: \beta_L < 0$, respectively. Since early and late treatment effects are independent, the null hypothesis of $H_{oel}: H_{oe} \cap H_{ol}$ can be tested by combining two sub hypothesis [1].

III. Results and Discussion

3.1. Simulation Results

In order to evaluate the performance of extended Cox model for testing the effect of treatment in randomized clinical trial when proportional hazards assumption is in doubt, survival data was simulated from a population exhibiting different possible treatment effects as displayed in Figure 1-5 below.

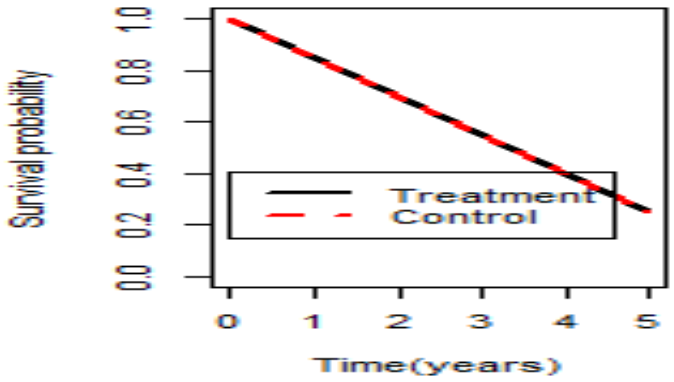

Figure 1. True survival curves under the null hypothesis of no treatment effect

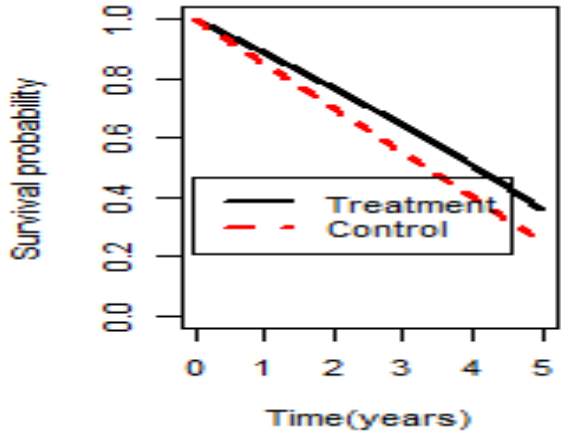

Figure 2: True survival curves under proportional hazards

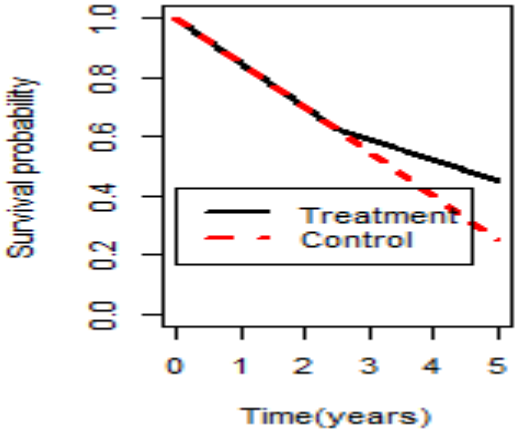

Figure 3: True survival curves under late treatment effect

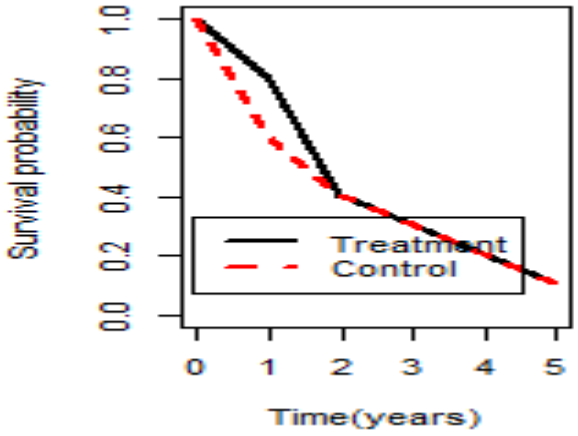

Figure 4: True survival curves under early treatment effect

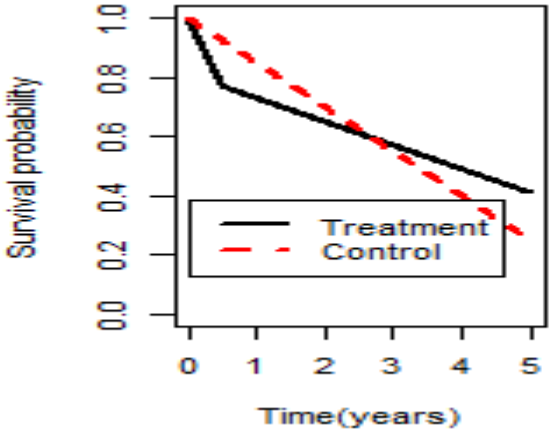

Figure 5: True survival curves under crossing survival curves

Figures 1-5 display true survival curves that were used to simulate sample data under five possible treatment effects. Specifically Figure 1 shows true survival curves under the null hypothesis of no treatment effect, Figure 2 shows true survival curves under constant beneficial treatment effect (i.e., proportional hazards), Figure 3 displays late treatment effect i.e., no initial effect of treatment but a gradually increasing beneficial effect, Figure 4 shows true survival curves under early treatment effect i.e., an initial beneficial effect of treatment that diminishes long-term and Figure 5 shows true survival curves that exhibits an initial harmful and late beneficial effect of treatment i.e., crossing survival curves. The sample data was replicated 1000 times under each scenario.

Table 1. Estimated type I error rates and Powers of the test based on the simulations under different scenarios from Extended Cox model without including covariate.

		Percentage of administrative censoring		
scenarios	Sample size per group	**10**%	**30**%	**70**%
Null Hypothesis	50	0.054	0.057	0.073
	100	0.053	0.054	0.062
	200	0.051	0.053	0.056
Proportional hazards	50	0.501	0.450	0.329
	100	0.640	0.520	0.369
	200	0.643	0.579	0.394
Late Treatment	50	0.510	0.500	0.410
	100	0.552	0.522	0.430
	200	0.652	0.575	0.453
Early Treatment	50	0.566	0.590	0.706
	100	0.620	0.890	0.902
	200	0.690	0.904	0.934
Crossing Survivals	50	0.750	0.820	0.605
	100	0.840	0.870	0.620
	200	0.923	0.903	0.760

Table 1 displays simulation results of the estimated type I error rates and powers of test from extended Cox model without including the covariate in the model. To investigate type I errors from the model, the effect of treatment was considered under the null hypothesis of no treatment effect (Figure 1) and various censoring rates were also considered. Hence, it can be observed that a test from extended Cox model controlled a type I error rate stabilizing around the targeted 0.05 level of significance regardless of sample sizes and 10% and 30% censoring rates. This was expected in order for the test method to be efficient. However, the type I error rates are slightly inflated at 70% censoring rate although it gradually decrease as the sample size increase under each treatment group (see Figure 6 also).

The powers of the test from the model are expected to depend heavily on the scenarios, for instance, under proportional hazards alternative (Figure 2), the test had moderate power around 60% at 10% censoring rate regardless of sample size considered in the study and decreased as censoring rate increases from 10% to 70%. However, at 70% censoring rate, the power of test becomes less than 40% which is a relatively low level (Table 1, Figure 7). In the case of late treatment effect (Figure 3), the test from the model was moderate with power of the test about 50% at censoring rates of 10% and 30% for all combinations of sample size per treatment group considered in the study and the power of the test slightly increase with increasing sample size per treatment group from 50 to 200. By contrast, under this scenario the powers of the test gradually decrease to with increasing censoring rates regardless of sample size per treatment group (Table 1, Figure 8).

In the situation where two survival curves are separate at the beginning and then close as time goes on (i.e., early

treatment effect,(Figure 4), test demonstrate powers above 80% when sample size of 100 and 200 at censoring rate of 30% and 70% and power of the test is less at 10% censoring rate. It was also noted that powers of the test was gradually improved increase in censoring rates and sample size per group. When the sample size of each group is 50, the power of the test is less regardless of censoring rate compared to sample size of 100 and 200 for each treatment group. Under this scenario, the test had better performance to detect early treatment effect in which hazards are expected to cross as sample size increase (Table 1, Figure 9).

As it was expected under crossing survival curves(Figure 5), the test for treatment effect from extended Cox model had better performance with power above 80% under 10% and 30% censoring rate for all combination of sample size under each treatment group. However, it was noted that the power of test gradually decrease to 60% with increasing censoring rate to 70%. Even if 70% of censoring rate appeared to influence the power of test, test from extended Cox model exhibits the optimal power in this situation (Table 1, Figure 10).

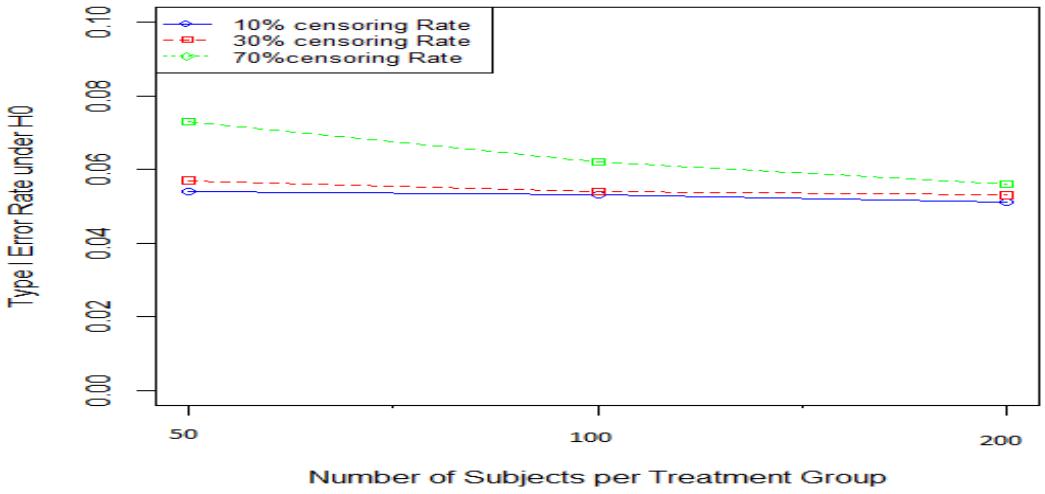

Figure 6: Estimated type I error rates of test from extended Cox model for variety of censoring rates

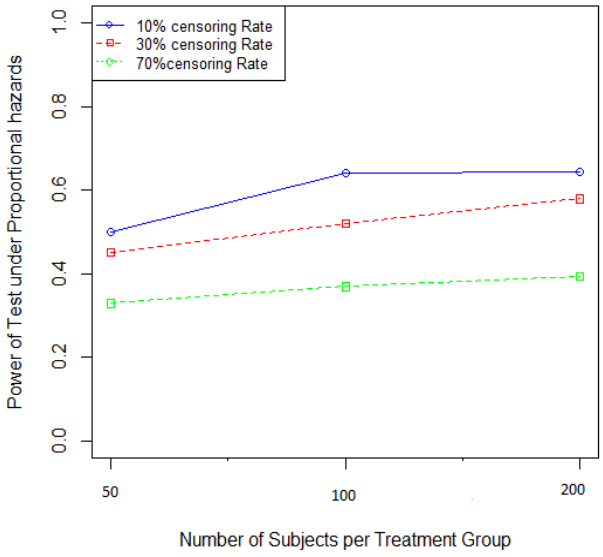

Figure 7: Powers of test under proportional Hazards

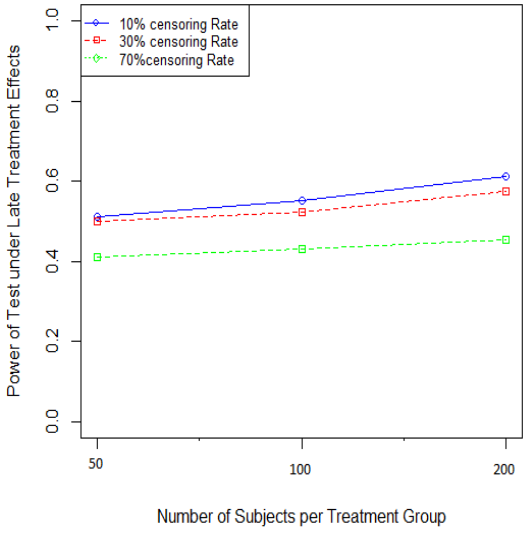

Figure 8: Powers of test under late treatment effect

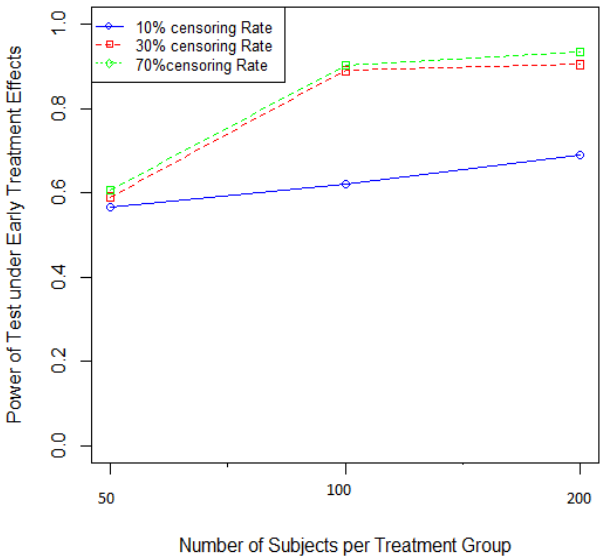

Figure 9: Powers of test under early treatment effects curves

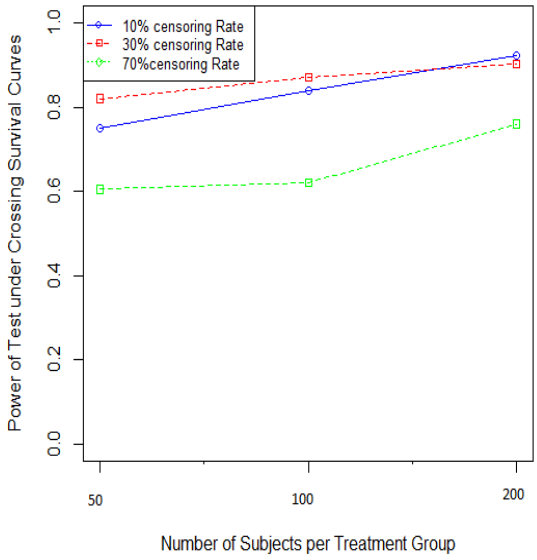

Figure 10: Powers of test under crossing survival

Table 2. Estimated type I error rates and Powers of the test based on the simulations under different scenarios from Extended Cox model with including covariate.

		Percentage of administrative censoring		
scenarios	Sample size per group	**10**%	**30**%	**70**%
Null Hypothesis	50	0.050	0.053	0.083
	100	0.052	0.050	0.072
	200	0.051	0.049	0.066
Proportional hazards	50	0.505	0.420	0.308
	100	0.580	0.440	0.323
	200	0.593	0.482	0.353
Late Treatment	50	0.610	0.590	0.321
	100	0.660	0.594	0.470
	200	0.636	0.658	0.453
Early Treatment	50	0.766	0.790	0.806
	100	0.820	0.810	0.836
	200	0.902	0.864	0.874
Crossing Survivals	50	0.726	0.810	0.404
	100	0.811	0.840	0.609
	200	0.890	0.901	0.730

Table 2 displays simulation results of the estimated type I error rates and powers of the test from extended Cox model with the presence of covariate in the model. Estimated type I error rates of the test from the model under various censoring rates are presented in Table 2(first three rows). Simulation result revealed that at 10% and 30% censoring rates, the type I error rates for the test are gradually approached the nominal level of 0.05 regardless of sample sizes considered under treatment group. However, at 70% of censoring rate, the type I error rate of the test exceeds the nominal level of 0.05 (Table 2, first three rows).

Regarding the power of test, when two survival curves have proportional hazards, the powers of test rises with increase in sample size per treatment group although the

powers of the test are relatively low. By contrast, for a particular sample size, the powers of the tests decline with increasing censoring rate. Moreover, in the situation when there is late treatment effect the test exhibit similar pattern to that of under proportional hazards and remaining relatively low. There was a slight loss in power under both scenarios when covariate is included in the model.

On the other hand, when there is early treatment effect, the test exhibits powers of 80% and above regardless of the censoring rates for a sample size of 100 and 200 under each treatment group. Surprisingly, for increasing censoring rates, test exhibits gradually increasing power for all combination of sample size; however, power of the test is raised with increase in sample size under each treatment. For crossing of the survival curves, the test demonstrate powers greater than 80% regardless of sample size under each treatment group at censoring rate of 10% and 30%. However, for increasing censoring rates, the test reveals gradually decreasing powers. When sample sizes are 100 and 200 under each treatment group, the test maintained powers of 80% and above at censoring rates of 10% and 30% when covariate is introduced in the model. When sample size is equal to 50 and censoring rate is 70%, the power of test becomes about 40% which is a relatively low level.

3.2. Discussion

In many cancer clinical trials, which require long-term follow up, a new treatment group is compared with an existing standard treatment or control group based on the hazard ratio estimated by the Cox model in which the hazard ratio often changes over time[14]. In such cases, the test for the difference in hazard ratio between the treatments will most likely not be significant and Cox test is known to lose power, because of the contrasting early and late effects of the treatments; furthermore it is difficult to interpret the hazard ratio estimated by the Cox model. Hence, different approaches have been proposed as alternative to the Cox proportional hazards model in the case of non-proportional hazards and extended Cox model is advocated for survival data analysis in such a case when there is long-term follow [14],[19-20],[23]. However, most studies have not evaluated the statistical properties of test from extended Cox model under varieties of censoring rates specifically at high censoring rate. Therefore, the main purpose of this study was to evaluate the performance of extended Cox model for testing the treatment effect in randomized clinical trials when proportional hazards assumption is not satisfied with variations in censoring rates. This was done based on data from simulation and performance of test was evaluated in terms of maintaining nominal level of significance and empirical power. From a simulation result, it was seen that the test from extended Cox model controlled the type I error rate accurately at the targeted 0.05 level of significance with and without including covariate in the model at 10% and 30% censoring rates regardless of sample sizes of 50,100 and 200 per treatment group. This result was in line with findings by Andrea Callegaro and Bart Spiessens [14]. However, at 70% of censoring rate, the estimated type I error rates of the test exceeds the nominal level of 0.05 regardless of sample sizes considered in the study.

It was observed that the performance of the test from extended model for testing treatment effect generally lacks power in situations where there is late treatment effect and two survival curves proportional. This result was also in line with findings by Andrea Callegaro and Bart Spiessens [14]. Simulation result showed that the extended Cox model in testing treatment effect performed reasonably well with power of about 80% under early treatment effect where hazards are expected to cross at sample sizes of 100 and 200 per treatment group regardless of censoring rates. In addition, when censoring rate was increased from 10% to 70%, the powers of test were improved under early treatment effect where hazards are expected to cross. Such an increase in powers might be due to the fact that increasing in censoring rates from 10% to 70% may lead to a shortening of the follow-up period.

Moreover, the simulation results also suggested that the powers of test statistic rise with increasing sample size at 10% and 30% censoring rates under crossing survival curves. By contrast, the powers of the test gradually decline at censoring rate of 70%. When sample size is equal to 50 per treatment group and when the censoring rate is increased to 70%, the power of each test becomes about 40%, a relatively low level. However, when sample sizes are 100 and 200 per treatment group while the censoring rates are 10% and 30% test demonstrates powers above 80%. As it was expected, the test from extend Cox model showed the optimal power in this situation at 10% and 30% of censoring rates. It should be noted that the performance of test from extended Cox depends on choice for heavy side function for time t. Hence, the choice for heavy side function must be pre-specified in the protocol.

IV. CONCLUSIONS

The main aim of this paper was to evaluate the performance of the extended Cox model in order to test treatment effect in randomized clinical trials with possible

non proportional hazards with variation of censoring rates under 50,100,200 subjects per treatment group. Hence from simulation results, it can be concluded that the test statistic from extend Cox model had a robust power with reasonable type I error rates at censoring rate of 10% and 30% under situation of crossing survival curves and the model allows adjusting for baseline covariates as it performed reasonable well with the presence of baseline covariate in the model. However, test exhibits relatively low powers and estimated type I error rates of the test were slightly inflated from the nominal level of 0.05 at censoring rate of 70% regardless of sample sizes considered in the study. It is duly acknowledged that this paper does not cover in-depth about extended Cox model for survival analysis and evaluated the method based on real data set; this might require much future research. Nonetheless, the paper can serve as a good note for the beginners who are interested to learn about the behavior of extended Cox model with variations in censoring rates in the situation where proportional hazards assumption is not satisfied.

ACKNOWLEDGEMENTS

The author thanks Andrea Callegaro (Ph.D.) from GSK Pharmaceuticals Company and Prof. dr. Roel BRAEKERS from Hasselt University, Belgium for all their constructive comments and professional support during writing of this paper. He also thanks Hasselt University for providing necessary facilities to conduct my master thesis through which this piece of work was carried out. Furthermore, the author thanks Dr. Chandra Sekarreddy from Arba Minch University, statistics department for analytically reviewing the paper.

AVAILABILITY OF DATA

The R code used in simulations and supporting conclusions of this article is available by contacting author.

AUTHORS' CONTRIBUTIONS

BBA designed the study, drafted the manuscript and reviewed the article. Author read and approved the final manuscript.

ETHICAL CONSIDERATIONS

The study was carried out after getting permission from the ethical clearance committee of Hasselt University. After Ethical clearance and approval of the University Ethical Committee the actual research activities was undertaken.

COMPETING INTERESTS

The author declares that he has no competing interests.

REFERENCES

[1] Fisher, R. A. (1925). Statistical Methods for Research Workers. Oliver and Boyd (Edinburgh). ISBN 0-05-002170-2.

[2] Yang, S and Zhao, Y. (2007). Testing treatment effect by combining weighted log-rank tests and using empirical likelihood: Science Direct: Statistics & Probability Letters 77: 1385–1393.

[3] Van Houwelingen, H.C, and Putter, H. (2014). Comparison of stopped Cox regression with direct methods such as pseudo-values and binomial regression. Lifetime Data Anal, DOI 10.1007/s10985-014-9299-3.

[4] Logan, B.R, Klein, J.P and Zhang, M.J. (2008). Comparing Treatments in the Presence of Crossing Survival Curves:An Application to Bone Marrow Transplantation: *Biometrics: 733–740. doi:10.1111/j.1541-0420.2007.00975.x.*

[5] Cox, D. R. (1972). Regression models and life-tables. Journal of the Royal Statistical Society, Series B, 34 (2), 187- 220.

[6] Putter, H, Sasako, M, Hartgrink, H. H, van de, Velde C. J and van Houwelingen, J. C. (2005). Long-term survival with non-proportional hazards: results from the Dutch Gastric Cancer Trial. Stat Med, 24, 2807-2821.

[7] Royston, P. and Parmar, M. K. B. (2011). The use of restricted mean survival time to estimate the treatment effect in randomized clinical trials when the proportional hazards assumption is in doubt. Statist. Med., 30, 2409-2421.

[8] Schemper, M., Wakounig, S., and Heinze, G. (2009). The estimation of average hazard ratios by weighted Cox regression. Statist. Med., 28, 2473-2489.

[9] Klein, J.P, Logan, B.R and Harhoff, M and Andersens, P. K. (2007). Analyzing survival curves at a fixed point in time. Statistics in Medicine, 26, 4505-4519.

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

[11] Chen, P and Tsiatis, A. A. (2001). Causal inference on the difference of the restricted mean lifetime between two groups, Biometrics vol. 57 pp. 1030–1038.

[12] Machin D, Cheung YB, Parmar MK..,(2006). Survival analysis a practical approach, 2nd edition, West Sussex: John Wiley and Sons Ltd.

[13] Leung KM, Elashoff RM, Afifi AA.(1997). Censoring issues in survival analysis. Annu Rev Public Health. 1997;18:83–104.

[14] Andrea Callegaro & Bart Spiessens (2017) Testing Treatment Effect in Randomized Clinical Trials With Possible Nonproportional Hazards, Statistics in Biopharmaceutical Research, 9:2, 204-211, DOI:10.1080/19466315.2016.1257436.

[15] Burton, A, Altman, D.G, Royston, P, and Holder, R.L. (2006). The design of simulation studies in medical statistics: Wiley InterScience, *Statist. Med*, 25:4279–4292.

[16] Lee, S. H. (2007), "On the Versatility of the Combination of the Weighted Log-Rank Statistics," *Computational Statistics & Data Analysis*, 51,6557–6564.

[17] Campbell, H., and Dean, C. B., (2013). The Consequences of Proportional Hazards Based Model Selection, *Statistics in Medicine*, 33, 1042–1056.

[18] Li,H.,Han,D.,Hou, Y.,Chen,H., andChen, Z. (2015). Statistical Inference Methods for Two Crossing Survival Curves: A Comparison of Methods, *PLoS One*, 10, e0116774.

[19] Pettitt, A. N. and Daud, I. Bin (1990). Investigating time dependence in Cox's proportional hazards model, Applied Statistics 39, 313-329.

[20] Hartina Husain , Sri Astuti Thamrin, Sulaiha Tahir , Ahmad Mukhlisin , Mirna Apriani M. (2018). The Application of Extended Cox Proportional Hazard Method for Estimating Survival Time of Breast Cancer; Journal of Physics: Conf. Series **979** (2018) 012087 doi :10.1088/1742-6596/979/1/012087.

[21] Gillen, D. L and Emerson, S. S. (2005). A note on P-Values under Group Sequential Testing and Non proportional Hazards: Biometrics 61, 546-551, *DOI: 10.1111/j.1541-0420.2005.040342.x.*

[22] Liang, K.Y and Zeger, S.L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, 73, 13-22.

[23] Fisher, L. D. ve Lin D. Y. (1999). Time-dependent covariates in the Cox proportional hazards regression model, Annual Review of Public Health 20, 145-157.