Research Article

Statistics
in Medicine

Received XXXX

(www.interscience.wiley.com) DOI: 10.1002/sim.0000

Supplementary Appendix: Subgroup Detection and Sample Size Calculation with Proportional Hazards Regression for Survival Data

Suhyun Kang, Wenbin Lu*† and Rui Song

1. Additional simulation results for subgroup identification

We reported here the sensitivity and specificity of our proposed method for identifying the subgroup under the change-plane model considered in Section 4.1.2. of the paper. The average size of identified subgroups are also given. The true subgroup size is 520. The results are given in Table 1. As the magnitude of treatment effect increases, sensitivity and specificity increase, and the estimated subgroup size becomes closer to the true value.

Table 1. Sensitivity, specificity and average size for subgroup identification.

Treatment effect		B1	B2	B3
$\eta = 0.2$	Sensitivity	0.521	0.511	0.520
	Specificity	0.814	0.814	0.805
	Size of subgroup	354	349	358
$\eta = -0.2$	Sensitivity	0.508	0.513	0.520
	Specificity	0.808	0.814	0.800
	Size of subgroup	350	350	360
$\eta = 0.5$	Sensitivity	0.911	0.909	0.909
	Specificity	0.886	0.884	0.875
	Size of subgroup	512	513	517
$\eta = -0.5$	Sensitivity	0.898	0.900	0.905
	Specificity	0.879	0.873	0.870
	Size of subgroup	510	514	518
$\eta = 0.8$	Sensitivity	0.971	0.972	0.971
	Specificity	0.938	0.935	0.930
	Size of subgroup	517	519	520
$\eta = -0.8$	Sensitivity	0.954	0.955	0.959
	Specificity	0.924	0.925	0.923
	Size of subgroup	516	515	518

Department of Statistics, North Carolina State University, Raleigh, NC 27695, U.S.A.
** Correspondence to: Wenbin Lu, Department of Statistics, North Carolina State University, Raleigh, NC 27695, U.S.A.*
† E-mail: lu@stat.ncsu.edu

2. Simulation results with censoring rate of 75%

We have conducted additional simulations for scenarios where the censoring rate is 75% and the baseline effect model is linear. Other settings are the same as those in Sections 4.1.1 and 4.1.2. of the paper. The results for $N = 1000$, $N = 2000$ and $N = 3000$ based on 500 runs are given in Table 2.

Table 2. Simulation results for 75% censoring rate.

		Type I error			Power	
		$\alpha = 0.05$	$\alpha = 0.1$		$\alpha = 0.05$	$\alpha = 0.1$
$N = 1000$	$\eta = 0$	0.03	0.07	$\eta = 0.5$	0.656	0.752
				$\eta = -0.5$	0.664	0.786
$N = 2000$	$\eta = 0$	0.03	0.07	$\eta = 0.5$	0.960	0.982
				$\eta = -0.5$	0.940	0.962
$N = 3000$	$\eta = 0$	0.05	0.09	$\eta = 0.5$	1	1
				$\eta = -0.5$	1	1

The results shows that when the censoring rate is 75%, the type I errors of the proposed test are slightly lower than the nominal level with the sample sizes $N = 1000$ and $N = 2000$. However, as the sample size increases to $N = 3000$, the type I errors are close to the nominal level. In addition, the power increases as the sample size increases.

3. Simulation results with $p = 4$ covariates

We have conducted additional simulations for the cases with $p = 4$ covariates. Specifically, we consider four independent covariates: $X_1 \sim Ber(0.5)$, $X_2 \sim U[-1, 1]$, $X_3 \sim N(1, 0.5^2)$, and $X_4 \sim N(0, 0.5^2)$. We set $\gamma_0 = (-0.17, -0.301, -0.67, 0.239, -0.612)$ and the true subgroup proportion is approximately 50%. We used a spherical transformation to generate $M = 50000$ grid points for the subgroup parameter γ. We consider the linear baseline covariate effect model. The results for $N = 1000$ and $N = 2000$ based on 500 runs are given in Table 3.

Table 3. Simulation results for cases with $p = 4$.

		Type I error			Power	
		$\alpha = 0.05$	$\alpha = 0.1$		$\alpha = 0.05$	$\alpha = 0.1$
$N = 1000$	$\eta = 0$	0.044	0.100	$\eta = 0.5$	0.928	0.966
				$\eta = -0.5$	0.926	0.960
$N = 2000$	$\eta = 0$	0.046	0.098	$\eta = 0.5$	1	1
				$\eta = -0.5$	1	1

Under all scenarios, the type I errors are close to the nominal level and the powers are comparable to those with two covariates. In addition, we report the average (in seconds) and standard deviation of the computation time for different numbers of covariates, $p = 2$ and $p = 4$. We considered the linear baseline covariate effect model with sample size $N = 1000$ and 15% censoring rate. We used $M = 10000$ grid points of γ for $p = 2$, while used $M = 50000$ for $p = 4$. In addition, we used 1000 resamplings for both cases. As shown in Table 4, the computational time increases drastically as the number of covariates increases. However, it took less than one minute on average for one simulation with $p = 4$.

In general, the proposed method can work reasonably well for a small number of covariates but may be time-consuming when the number of covariates is large.

Table 4. Computational time in seconds

	mean	sd
$p = 2$	4.830	0.295
$p = 4$	43.700	0.975

4. Sample size studies when the true model is not from the change-plane model

We have conducted additional simulations for sample size and power calculation under the smooth treatment effect using the derived procedure based on the change-plane model. Here, we considered a single covariate which follows a uniform distribution on $[-1, 1]$. The survival times were generated from the following hazards model

$$\lambda(t|A_i, X_i) = \lambda(t)e^{X_i+\eta A_i \Phi(\frac{X_i-\gamma_0}{\sigma})},$$

where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal random variable and $\sigma = \text{sd}(X_i)$. In addition, we considered the censoring rate 15%. The values for the grid points of γ and the "true" change plane parameter are similarly obtained as in Section 4.2. for sample size calculation based on the change-plane model. In Table 5, we report the required sample size that gives 90% power at the 0.05 level of significance and the empirical power of the proposed test for detecting the subgroup with the estimated sample size.

Table 5. Power and sample size calculation for smoothed treatment effect

η	γ_0	Sample Size	Power
0.2	0.5	6923	0.88
	0	3370	0.89
	-0.5	2601	0.92
0.4	0.5	1811	0.89
	0	980	0.91
	-0.5	648	0.92

As expected, the required sample size increases as the treatment effect magnitude and the subgroup size decrease. In addition, under all scenarios, the empirical powers are close to the nominal level even when the true model is not from the change-plane model, showing certain degree of robustness of the proposed sample size formula to the misspecification of the change-plane model.

5. Simulations with a nonzero main effect of treatment

We have conducted some simulations with a nonzero main effect of treatment. Specifically, the survival times are generated from the proportional hazards model

$$\lambda(t|X_i) = \lambda_0(t)e^{\theta' X_i+0.2A_i+\eta A_i I(\gamma' \tilde{X}_i \geq 0)}.$$

We consider the censoring rate 15% and sample sizes $N = 1000$ and $N = 2000$. When fitting the null model, we include the main effect of treatment. The simulation results for $N = 1000$ and $N = 2000$ based on 500 runs are given in Table 6.

Table 6. Simulation results with a nonzero treatment main effect.

		Type I error			Power	
		$\alpha = 0.05$	$\alpha = 0.1$		$\alpha = 0.05$	$\alpha = 0.1$
$N = 1000$	$\eta = 0$	0.04	0.07	$\eta = 0.5$	0.600	0.664
				$\eta = -0.5$	0.624	0.714
$N = 2000$	$\eta = 0$	0.05	0.09	$\eta = 0.5$	0.908	0.936
				$\eta = -0.5$	0.932	0.948

It can be seen that the type I errors are close to the nominal levels especially when $N = 2000$ and the power looks comparable to the cases when the treatment main effect is not included.

6. Simulations for the misspecified proportional hazards model

We have conducted additional simulations under the proportional odds model. However, in our implementation, we still fit a proportional hazards model under the null. Specifically, we consider similar settings with the linear baseline effect and 15% censoring rate. The results for sample sizes $N = 1000$ and $N = 2000$ based on 500 runs are given in Table 7.

Table 7. Simulation results for the misspecified proportional hazards model.

		Type I error			Power	
		$\alpha = 0.05$	$\alpha = 0.1$		$\alpha = 0.05$	$\alpha = 0.1$
$N = 1000$	$\eta = 0$	0.03	0.06	$\eta = 0.5$	0.498	0.600
				$\eta = -0.5$	0.586	0.686
$N = 2000$	$\eta = 0$	0.04	0.07	$\eta = 0.5$	0.878	0.926
				$\eta = -0.5$	0.884	0.940

The results are comparable to those under the proportional hazards model reported in the paper, however, the powers are slightly lower than those under the proportional hazards model.

7. Simulations when censoring times depend on treatment

We have conducted additional simulations for scenarios where the censoring times are generated from the model with the hazard function $\lambda(t|X_i, A_i) = \lambda_{0c} exp(\theta_c' X_i + \tau_c A_i)$ and $\tau_c \neq 0$. The constant λ_{0c} was chosen to give the censoring rate of 15%. We consider the same setting with the linear baseline effect model as studied in Section 4.1. of the paper. The results for sample sizes $N = 1000$ and $N = 2000$ based on 500 runs are given in Table 8. Based on the results, although censoring times depend on treatment, our proposed test still gives reasonable performance. However, in general, the proposed test may not be valid when this assumption is violated.

Table 8. Simulation results when censoring times depend on treatment.

		Type I error			Power	
		$\alpha = 0.05$	$\alpha = 0.1$		$\alpha = 0.05$	$\alpha = 0.1$
$N = 1000$	$\eta = 0$	0.03	0.06	$\eta = 0.5$	0.982	0.992
				$\eta = -0.5$	0.976	0.986
$N = 2000$	$\eta = 0$	0.03	0.07	$\eta = 0.5$	1	1
				$\eta = -0.5$	1	1