Received: Added at production | Revised: Added at production | Accepted: Added at production

DOI: xxx/xxxx

ARTICLE TYPE

BST 222, Project 2, Recurrence of paper "Association of Remnant Liver Ischemia With Early Recurrence and Poor Survival After Liver Resection in Patients With Hepatocellular Carcinoma"[1]

Bohao Zou

[1]Statistic, UC Davis, CA, USA

Present Address
Davis, CA, USA

Summary

In this project, we need to simulate the data, which comes from the paper "*Association of Remnant Liver Ischemia With Early Recurrence and Poor Survival After Liver Resection in Patients With Hepatocellular Carcinoma*" and we need to recur the result by using some methods of survival analysis. The main objective of original paper is "*To evaluate whether remnant liver ischemia (RLI) may have an adverse effect on long-term survival and morbidity after liver resection in patients with hepatocellular carcinoma*". The main conclusion and relevance of the original paper is that "*Preventive management and technical refinements in hepatectomy are important to decrease the risk of RLI and to improve survival of patients with hepatocellular carcinoma*". The two key points of this project are 1) how to simulate data and 2) how to analysis the data.

KEYWORDS:
Survival analysis methods, Simulate data

1 | SIMULATE DATA

The first key point of this recurrence is how to simulate the original data. There are 20 covariates in this dataset. Two of them are continuous variable and the other are category variable. I simulated those variables one by one according with the descriptions from the original paper. If the authors did not mention the distribution or the probability of one variable, for category variable, I will give the same probabilities to each categories. For continuous variable, I will use the normal distribution with the predefined mean or standard deviation. I will search some relative information about this continuous variable and then define the mean and SD.

1.1 | Generate Covariates

There are 328 patients who underwent liver resection for HCC. The details of how to simulate the data are showed below:

1. *Sex:* From the original paper, 252 were male and 76 were female. I generated 252 male samples and 76 female samples, combined those samples as a vector and then shuffled the vector.(Male = 1)

[0]**Abbreviations:** RLI, remnant liver ischemia; HCC, Hepatocellular carcinoma

2. *Age:* In the original paper, the minimal age is 26 and the maximum age is 83. For generating the age variable, I created a sequence from 26 to 83 and than take sample from this sequence.

3. *Disease Free:* According to this paper, there were 241, around 73.5% patients who had predominant etiology of liver disease, hepatitis B. I took samples from the Bernaulli distribution, which the probability with disease is 0.735.

4. *RLI type:* there are five RLI type in this dataset, its are none, marginal, partial, segmental and necrotic and the corresponding samples number are 124, 106, 63, 16, 19 respectively. The method that I used to generate this variable is very like the approach that I generated sex variable. Generated corresponding number of samples, combined them and shuffle the vector.

5. *Minimal or Sever RLI:* Based on those grades, patients with none or marginal RLI type were all included in the Minimal RLI class. The other patients were included in Sever RLI class. (Sever = 1)

6. *ICGR at 15 min > 10 %*: It is a binary category variable. In this paper, the authors did not depict this variable. I sampled from Bernaulli distribution with 0.5 probability. (Yes = 1)

7. *History of transarterial chemoembolization(TACE):* It is a binary category variable. In this paper, the authors did not give any information about this variable. I sampled from Bernaulli distribution with 0.5 probability. (Have TACE = 1)

8. *Child-Pugh classification B:* It is a binary category variable. In this paper, the authors did not give any information about this variable. I sampled from Bernaulli distribution with 0.5 probability (B = 1).

9. *Use of the Pringle maneuver:* It is a binary category variable. In this paper, the authors did not give any information about this variable. I sampled from Bernaulli distribution with 0.5 probability (Used = 1).

10. *Intraoperative transfusion:* It is a binary category variable. In this paper, the authors did not give any information about this variable. I sampled from Bernaulli distribution with 0.5 probability (Yes = 1).

11. *Serum albumin level:* It is a continuous variable. I searched on the internet. A normal albumin range is 3.4 to 5.4 g/dL. By using this information, I generated this variable by using normal distribution with mean equaled with 4.4 and standard deviation equaled with 3.

12. *Serum ALT level:* It is a continuous variable. From the internet, A normal ALT level is 29 to 33 units per liter (IU/L) for males and 19 to 25 IU/L for females. I used the Normal distribution with mean equaled with 26 and standard deviation equaled with 7 to generate this data.

13. *Longer operative time:* It is a binary category variable. In this paper, the authors did not give any information about this variable. I sampled from Bernaulli distribution with 0.5 probability (Long time = 1).

14. *Stage T3 or T4:* It is a binary category variable. In this paper, the authors did not give any information about this variable. I sampled from Bernaulli distribution with 0.5 probability (T3 = 1).

15. *Open surgery:* It is a binary category variable. In this paper, the authors did not give any information about this variable. I sampled from Bernaulli distribution with 0.5 probability (Open = 1).

16. *Nonanatomical resection :* It is a binary variable. It depicts if the patient had a nonanatomical resection. Because the authors did not give any description about this variable. I sampled from Bernaulli distribution with 0.5 probability (NonAnatomical = 1).

17. *Presence of a satellite nodule:* It is a binary variable. Because the authors did not give any description about this variable. I sampled from Bernaulli distribution with 0.5 probability (Exist = 1).

18. *Microscopic vascular invasion:* It is a binary variable. If one patient had microscopic vascular invasion. Because the authors did not give any description about this variable. I sampled from Bernaulli distribution with 0.5 probability (invasion = 1).

19. *Multinodular confluent or infiltrative gross tumor type:* It is a binary variable. In this paper, the authors did not give any information about this variable. I sampled from Bernaulli distribution with 0.5 probability (confluen = 1).

20. *Histologically confirmed cirrhosis:* It is a binary variable. In this paper, the authors did not give any information about this variable. I sampled from Bernaulli distribution with 0.5 probability (Confirmed = 1).

1.2 | Generate Time & Censoring

In this paper, the authors gave three different multivariable analysis tables in different situations. They analyzed the risk factors for *Severe RLI after Hepatectomy* first. Then they analyzed the risk factors for *Overall patients* and *Disease Free patients*. Some variables in the result table are different. For generating *time* variable from weibull distribution, we need to use the $log(HazardRatio)$ as the parameters to build this distribution. However, some variables are different in those three table. So, I defined a rule of how to use the $HazardRatio$ coefficients to build the weibull distribution.

1. If one variable has appeared in the result table of *Overall patients*, I will use this $HazardRatio$.
2. If one variable doesn't appear in the table of *Overall patients* but it appears in the table of *Disease Free patients*, I will use the $HazardRatio$ from the table of *Disease Free patients*.
3. If one variable only appears in the table of *Severe RLI after Hepatectomy*, I will use this $HazardRatio$.

By using this rule and setting the $scale$ parameter of $rweibull$ function as

$$scale = \vec{\beta}^T \vec{Z} = \beta_1 \times Z_1 + \beta_2 \times Z_2 + \cdots + \beta_p \times Z_p$$

where p represent the number of covariates and $p = 20$ in this project; β is the $log(HazardRatio)$ of each covariates, and Z is the covariates value, we can generate *time* variable.

In the original paper, the time range is 1-125 months. For generating the censoring variable, if time surpasses 125, it will have 50% probability to treated as censoring data. If the time between 1 and 125, it will have 20% probability to treated as censoring data.

2 | SURVIVAL ANALYSIS

In the next sections, I will use Cox-PH model to analysis the simulated data. This is because the objection of the original paper is "To evaluate whether remnant liver ischemia (RLI) may have an adverse effect on long-term survival and also wants to know which covariates can increase the risk of death for the patients who underwent liver resection with hepatocellular carcinoma". The Cox-PH model is the best method to achieve this objection.

For we can exactly recur the conclusion of original paper, in the processing of analysis, if one variable has NA value in the multivariable analysis table or not appears in the multivariable analysis table, I would not append this variable into the Cox-PH model for analysis. This procedure is similar with the processing of model selection. The only different is we only used the pre-specified variables in paper.

2.1 | Risk Factors for Severe RLI After Hepatectomy

From the paper, the authors analyzed the risk factors for severe RLI after hepatectomy first. In this project, I also followed with the mind of this paper and analyze the hazard ratio for severe RLI in the beginning. The result of original paper is showed below:

Table 1. Multivariable Analysis of the Risk Factors for Remnant Liver Ischemia After Hepatectomy

Variable	OR (95% CI)	*P* Value
Previous TACE	1.77 (1.03-3.04)	.04
Child-Pugh classification B or C	1.57 (0.57-4.35)	.38
Use of the Pringle maneuver	1.96 (1.08-3.58)	.03
Intraoperative transfusion	1.11 (0.60-2.06)	.75
Serum albumin level	0.75 (0.36-1.57)	.45
Serum ALT level	1.004 (0.998-1.011)	.20
Longer operative time	1.003 (1.002-1.005)	<.001

Abbreviations: ALT, alanine aminotransferase; OR, odds ratio; TACE, transarterial chemoembolization.

FIGURE 1 The result of severe RLI from original paper.

Except the Serum albumin level and Serum ALT level are continuous variables, the others are all binary category variables. From this compare, we can know that in the original paper, the multivariable analysis(Cox-PH model) showed the history of TACE, use of the Pringle maneuver and longer operative time were the independent risk factor for severe RLI after hepatectomy(Under the significant level of 0.05).

Analysis of Maximum Likelihood Estimates

Parameter	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio
p_TACE	1	0.26036	0.22805	1.3035	0.2536	1.297
chile_class	1	0.10953	0.23413	0.2188	0.6399	1.116
pringle_maneuver	1	0.28342	0.23965	1.3987	0.2369	1.328
intra_trans	1	-0.22129	0.22414	0.9747	0.3235	0.801
serum_alb_level	1	0.20185	0.04250	22.5538	<.0001	1.224
serum_alt_level	1	-0.01752	0.02030	0.7446	0.3882	0.983
long_op_time	1	0.03434	0.22600	0.0231	0.8792	1.035

FIGURE 2 The result of severe RLI of simulated data.

In my simulated data, the high value of Serum albumin level is the significant risk factor for the survival of sever RLI patients who underwent hepatectomy.(Under the significant level of 0.05). If we do not see the P-Value of my result, we can conclude that the history of TACE, use of the Pringle maneuver and longer operative time were the independent risk factor. We can roughly say that we have recurred the result of original paper. Because our data came from simulation and if the authors did not mention the distribution of some variable, we only take samples from Bernaulli distribution with equaled probabilities for each class. Thus, we can't guarantee the all variable are significant under the analysis of simulated data.

2.1.1 | Model Diagnostics for Sever Data

In this section, we should do diagnostic for the model. But this section did not appear in the original paper. We need to complete the all procedure of survival analysis. In the previous section, we have done the model selection and model interpretation. The first thing we need to do for diagnostic is to check the overall fitting of the model that fitted with sever data. We can use Cox-Snell residuals to check the overall fitting of the model. The Cox-Snell residual vs. the Estimated Cumulative Hazard Rate plot is showed below:

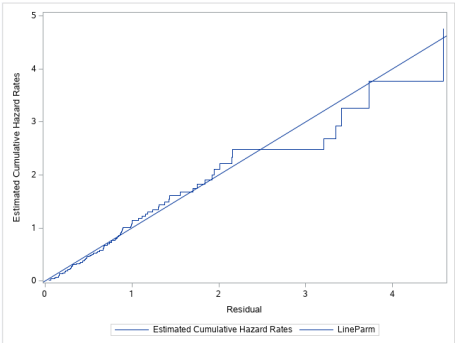

FIGURE 3 The Cox-Snell residual vs. the Estimated Cumulative Hazard Rate plot for sever.

From this plot we can know the model fits with the data very well. The second step we should do is to use Martingale residuals to check non-linearity of continuous variables. In this model, there are only two continuous variables. Serum albumin level and Serum ALT level. The plot of Serum albumin level or Serum ALT level vs. Martingale residuals plots are showed below:

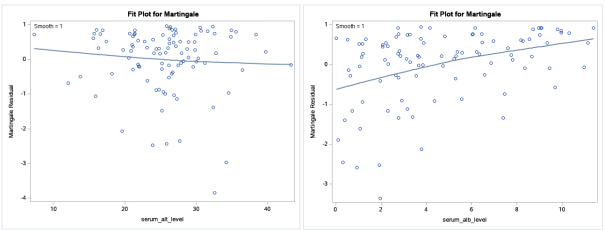

FIGURE 4 The plot of Serum albumin level and Serum ALT level vs. Martingale residuals. The left is Serum ALT level vs. Martingale residuals. The right is Serum albumin level vs. Martingale residuals

From this plot, we can know that those two variables are all linear in this Cox-PH model. The last thing we need to do is to check the PH assumptions for all covariates. We should use Schoenfeld residuals to test PH assumption. The ZPH correlation tests for Schoenfeld residuals and time are showed below:

zph Tests for Nonproportional Hazards

Transform	Predictor Variable	Correlation	ChiSquare	Pr > ChiSquare	t Value	Pr > |t|
RANK	p_TACE	-0.00066	0.000039	0.9950	-0.01	0.9951
RANK	chile_class	-0.0356	0.1279	0.7206	-0.33	0.7416
RANK	pringle_maneuver	-0.0752	0.5816	0.4457	-0.70	0.4861
RANK	intra_trans	0.00861	0.00684	0.9341	0.08	0.9365
RANK	serum_alb_level	-0.0388	0.1509	0.6976	-0.36	0.7200
RANK	serum_alt_level	-0.0180	0.0382	0.8451	-0.17	0.8680
RANK	long_op_time	-0.0761	0.5512	0.4578	-0.71	0.4810
RANK	_Global_	.	1.6162	0.9780	.	.

FIGURE 5 The result of ZPH tests

From those tests for covariates, we can know no variable violates the PH assumptions. This diagnostics indicates that we have a excellent survival analysis.

2.2 | Risk Factors for Overall Survival After Hepatectomy

From the paper, the authors analyzed the risk factors for overall patients after hepatectomy. Follow with the step of paper, in this section, we need to analyze the risk factors for overall patients by using our simulated data. The results are showed below:

Table 2. Multivariable Analysis of the Prognostic Factors Associated With Overall Survival and Disease-Free Survival

	Overall Survival		Disease-Free Survival	
Variable	OR (95% CI)	*P* Value	OR (95% CI)	*P* Value
Male sex	NA	NA	0.84 (0.57-1.25)	.39
ICGR at 15 min >10%	NA	NA	1.25 (0.90-1.73)	.19
Stage T3 or T4	1.66 (0.88-3.13)	.12	1.71 (1.07-2.72)	.03
Previous TACE	NA	NA	0.95 (0.67-1.35)	.78
Child-Pugh classification B or C	1.36 (0.77-2.42)	.29	NA	NA
Severe remnant liver ischemia	6.98 (4.27-11.43)	<.001	5.15 (3.62-7.35)	<.001
Open surgery	1.76 (1.10-2.82)	.02	NA	NA
Intraoperative transfusion	0.98 (0.61-1.58)	.95	NA	NA
Nonanatomical resection	NA	NA	1.57 (1.13-2.19)	.008
Presence of a satellite nodule	1.17 (0.63-2.19)	.62	0.80 (0.51-1.27)	.35
Microscopic vascular invasion	1.16 (0.50-1.42)	.51	1.60 (1.13-2.31)	.008
Multinodular confluent or infiltrative gross tumor type	2.76 (1.13-6.71)	.03	0.81 (0.34-1.94)	.64
Histologically confirmed cirrhosis	1.23 (0.74-2.04)	.43	NA	NA

FIGURE 6 The result of overall and disease free from original paper.

From above tables we can know that the severe RLI, open surgery and multinodular confluent or infiltrative gross tumor type were independent risk factor for overall survival. Other variables have roughly the same effect to the survival probability of patients.

Analysis of Maximum Likelihood Estimates

Parameter	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio
stage	1	0.17334	0.12727	1.8550	0.1732	1.189
chile_class	1	0.12214	0.12675	0.9285	0.3353	1.130
min_Severe_RLI	1	1.50777	0.14761	104.3354	<.0001	4.517
open_surgery	1	0.26478	0.12667	4.3693	0.0366	1.303
intra_trans	1	-0.04526	0.12862	0.1238	0.7250	0.956
satellite_nodule	1	0.07050	0.12889	0.2992	0.5844	1.073
micro_vascular_invas	1	0.24395	0.12879	3.5877	0.0582	1.276
mcig_type	1	0.70011	0.12964	29.1639	<.0001	2.014
h_confirmed_c	1	0.18969	0.12632	2.2550	0.1332	1.209

FIGURE 7 The result of overall for simulated data.

From the the result of simulated data(above table), we can know that the severe RLI, open surgery and multinodular confluent or infiltrative gross tumor type were all the independent risk factor for overall survival under the significant level of 0.05. The result of simulated data is as the same as the result from original paper. This phenomenon may indicate that we have a excellent data simulation procedure. It captures all the essential information from the original paper and we used the correct model and variables.

The authors also make a comparison of overall survival between Minimal or Sever RLI and all five RLI types. They used log-rank test and drew the K-M plot for evaluating whether RLI have adverse effect on survival. I also did the log-rank test and drew the K-M curves of survival probability to make a comparison with the original paper. The comparisons are showed below:

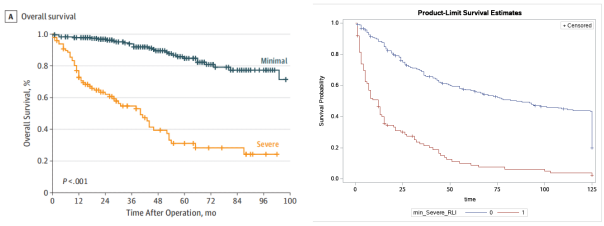

FIGURE 8 The comparison of K-M curve of minimum RLI or severe RLI between the original paper and the overall simulated data. The left plot comes from paper and the right plot comes from simulated data.

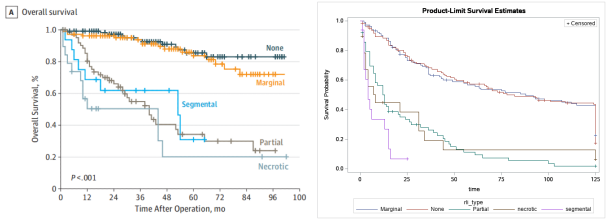

FIGURE 9 The comparison of K-M curve of all five RLI types between the original paper and the overall simulated data. The left plot comes from paper and the right plot comes from simulated data.

In the plots of original paper, the P-Value of log-rank tests are all smaller than 0.001. This shows that the survival probabilities have significant difference between different RLI types. In my simulated data, the statistical χ^2 are 110.4349 and 130.0566 for Minimal or Sever RLI types and all five different RLI types. This indicates that we have recurred the result of original paper and we have a very successful simulated dataset.

2.2.1 | Model Diagnostics for Overall Data

In this section, we should do diagnostic for the model that fitted with overall simulated data. But this section did not appear in the original paper. We need to complete the all procedure of survival analysis. The first thing we need to do for diagnostic is to check the overall fitting of the model that fitted with sever data. We can use Cox-Snell residuals to check the overall fitting of the model. The Cox-Snell residual vs. the Estimated Cumulative Hazard Rate plotis showed below:

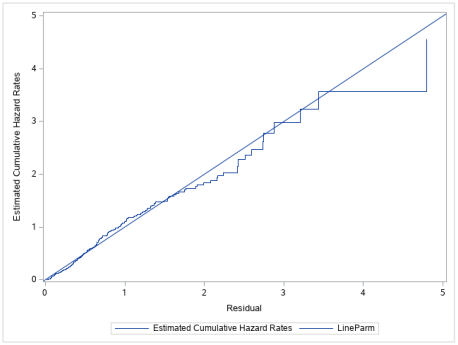

FIGURE 10 The Cox-Snell residual vs. the Estimated Cumulative Hazard Rate plot for overall.

From this plot we can know the model fits with the data very well. The second step we should do is to use Martingale residuals to check non-linearity of continuous variables. However, in this model, there is no continuous variable. We can ignore this step in

this section. The last thing we need to do is to check the PH assumptions for all covariates. We should use Schoenfeld residuals to test PH assumption. The ZPH correlation tests for Schoenfeld residuals and time are showed below:

zph Tests for Nonproportional Hazards

Transform	Predictor Variable	Correlation	ChiSquare	Pr > ChiSquare	t Value	Pr > |t|
RANK	stage	-0.0362	0.3313	0.5649	-0.58	0.5640
RANK	chile_class	-0.0995	2.5420	0.1109	-1.59	0.1122
RANK	min_Severe_RLI	-0.1737	7.0467	0.0079	-2.81	0.0053
RANK	open_surgery	-0.0607	0.9414	0.3319	-0.97	0.3333
RANK	intra_trans	-0.0560	0.8238	0.3641	-0.89	0.3725
RANK	satellite_nodule	-0.1156	3.5933	0.0580	-1.85	0.0648
RANK	micro_vascular_invasion	-0.0775	1.5981	0.2062	-1.24	0.2167
RANK	mcig_type	-0.1859	8.5763	0.0034	-3.02	0.0028
RANK	h_confirmed_c	-0.0522	0.6931	0.4051	-0.83	0.4056
RANK	_Global_	.	21.2726	0.0115	.	.

FIGURE 11 The result of ZPH tests.

From this table we can know that the variable Severe RLI and Multinodular confluent or infiltrative gross tumor type are not obey the PH assumptions. We can remove those variables if we want or use Stratified Cox PH Model to solve this problem.

2.3 | Risk Factors for Disease Free Survival After Hepatectomy

In the end, the authors analyzed the risk factors for disease free patients after hepatectomy. The main type of disease is Hepatitis B. Follow with the step of paper, in this section, we need to analyze the risk factors for disease free patients by using our simulated data. The result of simulated data is showed below:

Analysis of Maximum Likelihood Estimates

Parameter	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio
sex	1	-0.26892	0.27190	0.9782	0.3226	0.764
icgr_bigger_10	1	0.17051	0.24658	0.4782	0.4893	1.186
stage	1	0.07710	0.26609	0.0840	0.7720	1.080
p_TACE	1	-0.07733	0.25529	0.0918	0.7620	0.926
min_Severe_RLI	1	1.72442	0.31787	29.4298	<.0001	5.609
nonanatom_resection	1	0.64572	0.26077	6.1317	0.0133	1.907
satellite_nodule	1	0.05077	0.25458	0.0398	0.8419	1.052
micro_vascular_invas	1	0.06259	0.27309	0.0525	0.8187	1.065
mcig_type	1	0.72309	0.25805	7.8520	0.0051	2.061

FIGURE 12 The Cox-PH model result of simulated dataset.

In the original paper, the T stage of T3, severe RLI NAH resection and microscopic vascular invasion were independent risk factors for disease-free survival. In the result of simulated data, the severe RLI, NAH resection and Multinodular confluent or infiltrative gross tumor type are the independent risk factors for disease free survival under the significant of 0.05. If we do not see the P-Values of my result, we can get the roughly same conclusion of origianl paper.

Beside the Cox-PH model, the authors also used log-rank test to compare the hazard rates of minimum & severe RLI and to compare the hazard rates of five different RLI types. I also did the same steps as the authors. The comparison of my result and original paper are showed below:

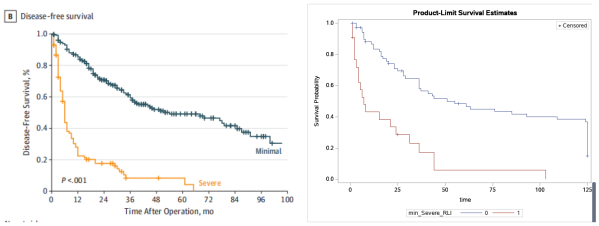

FIGURE 13 The comparison of K-M curve of minimum RLI or severe RLI between the original paper and the disease free simulated data. The left plot comes from paper and the right plot comes from simulated data.

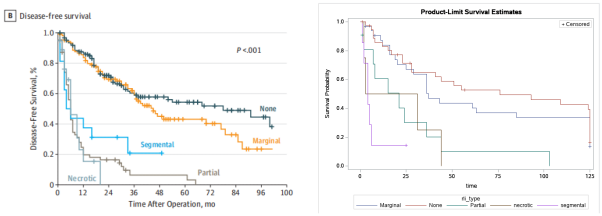

FIGURE 14 The comparison of K-M curve of all five RLI types between the original paper and the simulated disease free data. The left plot comes from paper and the right plot comes from simulated data.

The conclusions are the same as the analysis for overall survival. In the plots of original paper, the P-Value of log-rank tests are all smaller than 0.001. This shows that the survival probabilities have significant difference between different RLI types. In my simulated data, the statistical χ^2 are 26.6739 and 36.9953 for Minimal or Sever RLI types and all five different RLI types respectively. This indicates that we have recurred the result of original paper and we have a very successful simulated dataset.

2.3.1 | Model Diagnostics for Disease Free Data

In this section, we should do diagnostic for the model that fitted with disease free simulated data. But this section did not appear in the original paper. We need to complete the all procedure of survival analysis. The first thing we need to do for diagnostic is to check the overall fitting of the model that fitted with disease free data. We can use Cox-Snell residuals to check the overall fitting of the model. The Cox-Snell residual vs. the Estimated Cumulative Hazard Rate plot is showed below:

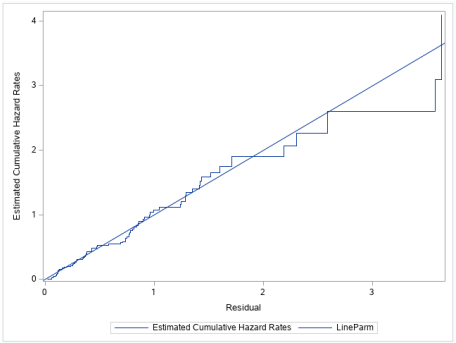

FIGURE 15 The Cox-Snell residual vs. the Estimated Cumulative Hazard Rate plot for disease free data.

From this plot we can know the model fits with the data very well. The second step we should do is to use Martingale residuals to check non-linearity of continuous variables. However, in this model, there is no continuous variable. We can ignore this step in

this section. The last thing we need to do is to check the PH assumptions for all covariates. We should use Schoenfeld residuals to test PH assumption. The ZPH correlation tests for Schoenfeld residuals and time are showed below:

zph Tests for Nonproportional Hazards

Transform	Predictor Variable	Correlation	ChiSquare	Pr > ChiSquare	t Value	Pr > |t|
RANK	sex	0.2277	3.7728	0.0521	1.97	0.0527
RANK	icgr_bigger_10	-0.0210	0.0362	0.8492	-0.18	0.8601
RANK	stage	-0.0948	0.7684	0.3807	-0.80	0.4252
RANK	p_TACE	0.0694	0.4135	0.5202	0.59	0.5598
RANK	min_Severe_RLI	-0.1043	0.7957	0.3724	-0.88	0.3799
RANK	nonanatom_resection	-0.1393	1.6906	0.1935	-1.19	0.2399
RANK	satellite_nodule	-0.0381	0.1182	0.7310	-0.32	0.7490
RANK	micro_vascular_invasion	-0.1088	1.0931	0.2958	-0.92	0.3594
RANK	mcig_type	-0.0463	0.1598	0.6894	-0.39	0.6972
RANK	_Global_	.	7.7362	0.5609	.	.

FIGURE 16 The result of ZPH tests.

From those tests for covariates, we can know no variable violates the PH assumptions. This diagnostics indicates that we have a excellent survival analysis

3 | CONCLUSION

By correctly simulating the data and using the correct model to analyze the dataset, we recurred the result of the original paper and the conclusions are roughly the same as the original paper. The severe RLI has adverse effect on the long-term survival and morbidity after liver resection in patients with hepatocellular carcinoma. In the patients with severe RLI, previous TACE, use the Pringle maneuver and longer operative time can increase the risk of death. The survival time of those patients could be improved by reducing the operative time and avoiding the Pringle maneuver. TACE before hepatectomy is not recommended. For overall patients, the severe RLI, open surgery and Multinodular confluent gross tumor type are the risk factors. We can improve the technology of surgery to prolong the survival time of patients. For disease free patients, a T stage of T3, severe RLI, NAH resection and microscopic vascular invasion are risk factors. A anatomical resection operation can fully clear the tumor. We can do a anatomical resection operation for those disease free patients to prolong their survival time.

References

1. Cho J, Han H, Choi Y, et al. Association of remnant liver ischemia with early recurrence and poor survival after liver resection in patients with hepatocellular carcinoma. *JAMA Surgery* 2017; 152(4): 386–392. Publisher Copyright: © 2017 American Medical Association. Copyright: Copyright 2017 Elsevier B.V., All rights reserved.doi: 10.1001/jamasurg.2016.5040

4 | APPENDIX

Codes for survival analysis (SAS):

```
proc import datafile = "/folders/myshortcuts/MyFolder/sever.csv" out=sever;

* analysis for sever RLI;
proc phreg data = sever;
model time*censor(0) = p_TACE chile_class pringle_maneuver
intra_trans serum_alb_level serum_alt_level long_op_time;
run;

* Cox-snell residuals plot for sever data;
proc phreg data = sever;
```

```
model time*censor(0) = p_TACE chile_class pringle_maneuver intra_trans serum_alb_level serum_alt_level
    long_op_time;
output out=plot1_1 logsurv=logsurv1 /method = ch;

data plot1_1;
set plot1_1;
snell = -logsurv1;
cons = 1;

proc phreg data=plot1_1;
model snell*censor(0) = cons;
output out = plot1_2 logsurv= logsurv2/method=ch;

data plot1_2;
set plot1_2;
cumhaz = - logsurv2;

proc sort data=plot1_2;
by snell;

proc sgplot data= plot1_2;
step y=cumhaz x=snell /MARKERFILLATTRS=(color="red");
lineparm x=0 y=0 slope=1; /** intercept, slope **/
label cumhaz = "Estimated Cumulative Hazard Rates";
label snell = "Residual";
run;

* PH assumption test for each variables in sever data;
ods noproctitle;
ods graphics / imagemap=on;

proc phreg data=sever zph(noplot global);
  model time*censor(0) = p_TACE chile_class pringle_maneuver intra_trans serum_alb_level serum_alt_level
    long_op_time / rl;
run;

/* ods noproctitle; */
/* ods graphics / imagemap=on; */
/*  */
/* proc phreg data=WORK.SEVER zph(noplot global) outest=Work.Phreg_est; */
/*  model time*censor(0)=p_TACE chile_class pringle_maneuver intra_trans  */
/*    serum_alb_level serum_alt_level long_op_time / rl; */
/*  output out=Work.Phreg_out resmart=resmart ressch=_all_; */
/* run; */

* Martingale Residuals for alt (continuous variable);
proc phreg data = sever;
model time*censor(0) = p_TACE chile_class pringle_maneuver
intra_trans serum_alb_level long_op_time;
output out=plot2_1 RESMART = Martingale;

proc loess data=plot2_1;
model Martingale = serum_alt_level / direct;
run;

* Martingale Residuals for alb (continuous variable);
proc phreg data = sever;
model time*censor(0) = p_TACE chile_class pringle_maneuver
intra_trans long_op_time serum_alt_level;
output out=plot2_1 RESMART = Martingale;

proc loess data=plot2_1;
model Martingale = serum_alb_level / direct;
run;


* analysis for overall RLI;
proc import datafile = "/folders/myshortcuts/MyFolder/overall.csv" out=overall;
proc phreg data = overall;
model time*censor(0) = stage chile_class min_Severe_RLI open_surgery
 intra_trans satellite_nodule micro_vascular_invasion mcig_type h_confirmed_c;
run;

*log-rank test for sever or min RLI for overall;
```

```
proc lifetest data=overall;
time time*censor(0);
strata min_Severe_RLI/ test=(logrank TARONE PETO MODPETO FLEMING(0,1));
run;


proc lifetest data=overall;
time time*censor(0);
strata rli_type/ test=(logrank TARONE PETO MODPETO FLEMING(0,1));
run;

* Cox-snell residuals plot for sever data;
proc phreg data = overall;
model time*censor(0) = stage chile_class min_Severe_RLI open_surgery
 intra_trans satellite_nodule micro_vascular_invasion mcig_type h_confirmed_c;
output out=plot1_1 logsurv=logsurv1 /method = ch;

data plot1_1;
set plot1_1;
snell = -logsurv1;
cons = 1;

proc phreg data=plot1_1;
model snell*censor(0) = cons;
output out = plot1_2 logsurv= logsurv2/method=ch;

data plot1_2;
set plot1_2;
cumhaz = - logsurv2;

proc sort data=plot1_2;
by snell;

proc sgplot data= plot1_2;
step y=cumhaz x=snell /MARKERFILLATTRS=(color="red");
lineparm x=0 y=0 slope=1; /** intercept, slope **/
label cumhaz = "Estimated Cumulative Hazard Rates";
label snell = "Residual";
run;

* PH assumption test for each variables in sever data;
ods noproctitle;
ods graphics / imagemap=on;

proc phreg data=overall zph(noplot global);
  model time*censor(0) = stage chile_class min_Severe_RLI open_surgery
 intra_trans satellite_nodule micro_vascular_invasion mcig_type h_confirmed_c / rl;
run;



* analysis for disease free RLI;
proc import datafile = "/folders/myshortcuts/MyFolder/diseaseFree.csv" out=diseaseFree;
proc phreg data = diseaseFree;
model time*censor(0) = sex icgr_bigger_10 stage p_TACE min_Severe_RLI
 nonanatom_resection satellite_nodule micro_vascular_invasion mcig_type;
run;



*log-rank test for sever or min RLI for overall;
proc lifetest data=diseaseFree;
time time*censor(0);
strata min_Severe_RLI/ test=(logrank TARONE PETO MODPETO FLEMING(0,1));
run;


proc lifetest data=diseaseFree;
time time*censor(0);
strata rli_type/ test=(logrank TARONE PETO MODPETO FLEMING(0,1));
run;


* Cox-snell residuals plot for sever data;
```

```
proc phreg data = diseaseFree;
model time*censor(0) = sex icgr_bigger_10 stage p_TACE min_Severe_RLI
 nonanatom_resection satellite_nodule micro_vascular_invasion mcig_type;
output out=plot1_1 logsurv=logsurv1 /method = ch;

data plot1_1;
set plot1_1;
snell = -logsurv1;
cons = 1;

proc phreg data=plot1_1;
model snell*censor(0) = cons;
output out = plot1_2 logsurv= logsurv2/method=ch;

data plot1_2;
set plot1_2;
cumhaz = - logsurv2;

proc sort data=plot1_2;
by snell;

proc sgplot data= plot1_2;
step y=cumhaz x=snell /MARKERFILLATTRS=(color="red");
lineparm x=0 y=0 slope=1; /** intercept, slope **/
label cumhaz = "Estimated Cumulative Hazard Rates";
label snell = "Residual";
run;

* PH assumption test for each variables in sever data;
ods noproctitle;
ods graphics / imagemap=on;

proc phreg data=diseaseFree zph(noplot global);
  model time*censor(0) = sex icgr_bigger_10 stage p_TACE min_Severe_RLI
 nonanatom_resection satellite_nodule micro_vascular_invasion mcig_type / rl;
run;
```

Codes for simulating dataset (R):

```
#########################################################################
# simulation if cox PH model is assumed, with some continuous covariates;
# baseline event time is assumed to follow Weibull/exponential distribution
# indepedent (uniform) censoring and Right censoring
#########################################################################
library(MASS)

sim_cox<- function(lambda0)
{
  # N = Total sample size
  # beta = PH coefficients : c()
  # lambda0 = rate parameter of the exponential distribution for baseline
  N = 328
  # sex male = 1, female = 0
  sex = sample(c(rep(1, 252), rep(0, 76)), size = N, replace = F)

  # age 26-83, mean 58,2
  age = sample(x = c(26:83), size = N, replace = T)

  # The predominant etiology of liver disease was hepatitis B
  liver_disease = sample(x=c("No Disease", "Hepatitis B"),
                         size=N, replace=TRUE, prob=c(1-0.735, 0.735))

  # RLI type
  rli_type = sample(c(sample("None", size = 124, replace = T),
                      sample("Marginal", size = 106, replace = T),
                      # severe RLI
                      sample("Partial", size = 63, replace = T),
                      sample("segmental", size = 16, replace = T),
                      sample("necrotic", size = 19, replace = T)), size = N, replace = F)
  # Minimal RLI or Severe RLI
  min_Severe_RLI = rep("Severe", N)
  min_Severe_RLI[which(rli_type == "None" | rli_type == "Marginal")] = "Min"

  # ICGR > 10%, yes = 1, no = 0
  icgr_bigger_10 =  sample(c(1, 0), size = N, replace = T)
```

```
# Previous TACE, tace = 1, notace = 0
p_TACE = sample(c(1, 0), size = N, replace = T)

# Child-Pugh classification B or C, b= 1, c=0
chile_class = sample(c(1, 0), size = N, replace = T)

# use of the Pringle maneuver
pringle_maneuver = sample(c(1, 0), size = N, replace = T)

# Intraoperative transfusion, yes = 1, no = 0
intra_trans = sample(c(1, 0), size = N, replace = T)

# Serum albumin level 3.4 to 5.4 g/dL
serum_alb_level = abs(rnorm(N, mean = 4.4, sd = 3))

# Serum ALT level 29 to 33 units per liter (IU/L) for males and 19 to 25 IU/L for females,
serum_alt_level = abs(rnorm(N, mean = 26, sd = 7))

# Longer operative time
long_op_time = sample(c(1, 0), size = N, replace = T)

# Stage T3 or T4, t3 = 1, t4 = 0
stage = sample(c(1, 0), size = N, replace = T)

# Open surgery
open_surgery = sample(c(1, 0), size = N, replace = T)

# Nonanatomical resection
nonanatom_resection = sample(c(1, 0), size = N, replace = T)

# Presence of a satellite nodule, exist = 1, noexist = 0
satellite_nodule = sample(c(1, 0), size = N, replace = T)

# Microscopic vascular invasion
micro_vascular_invasion = sample(c(1, 0), size = N, replace = T)

# Multinodular confluent or infiltrative gross tumor type, confluen = 1, infiltrative
mcig_type = sample(c(1, 0), size = N, replace = T)

# Histologically confirmed cirrhosis, yes = 1, no = 0
h_confirmed_c = sample(c(1, 0), size = N, replace = T)

#################################################################

# randomization to treatment or control
A = rep(1,N)
A[which(min_Severe_RLI == "Min")] = 0

# generate continuous covariates, mutually indepedent

# generate underlying event time (range, 1-125 months),
Time = rweibull(n=N, shape=1, scale = lambda0*exp(
1.943*(1-A) - 0.17*(1-sex) + 0.22 * (1-icgr_bigger_10) +
  0.5 * (1-stage) - 0.05 * (1-p_TACE) + 0.3*(1-chile_class)
+0.56*(1-open_surgery) - 0.02 * (1-intra_trans) +
  0.45 * (1-nonanatom_resection) + 0.15*(1-satellite_nodule)
+ 0.14 * (1-micro_vascular_invasion) + 1.01*(1-mcig_type)
+ 0.2*(1-h_confirmed_c)+0.67 * (1-pringle_maneuver)
-0.28*serum_alb_level + 0.0039*serum_alt_level + 0.002*(1-long_op_time))) + 1
time_new = c()
censor = c()
for (t in Time){
  if (t > 125){
    time_new = c(time_new, 125)
    if (runif(1)>0.5){
      censor = c(censor,0)
    }
    else{
      censor = c(censor,1)
    }
  }
  else{
    time_new = c(time_new, t)
```

```r
      if (runif(1) < 0.2){
        censor = c(censor,0)
      }
      else{
        censor = c(censor,1)
      }
    }
  }
  time_new = round(time_new)
  # data set
  dataset = data.frame(sex = sex, age = age,
  liver_disease = liver_disease, rli_type = rli_type,
                       min_Severe_RLI = A,
                       p_TACE = p_TACE, chile_class = chile_class,
                       pringle_maneuver = pringle_maneuver,
                       intra_trans = intra_trans,
                       serum_alb_level= serum_alb_level,
                       serum_alt_level= serum_alt_level,
                       long_op_time = long_op_time,
                       stage = stage, open_surgery = open_surgery,
                       nonanatom_resection = nonanatom_resection,
                       satellite_nodule = satellite_nodule,
                       micro_vascular_invasion = micro_vascular_invasion,
                       mcig_type= mcig_type,
                       h_confirmed_c=h_confirmed_c,
                       icgr_bigger_10 =
                       icgr_bigger_10,
                       time=time_new,
                       censor=censor)
  return(dataset)
}
mydata=sim_cox(lambda0 = 6)
write.csv(mydata, file="d:\\overall.csv", row.names = F)

### sever data
severData = mydata[which(mydata$min_Severe_RLI==1),]
write.csv(severData, file = "d:\\sever.csv", row.names = F)

### Disease free data
diseaseData = mydata[which(mydata$liver_disease=="No Disease"),]
write.csv(diseaseData, file = "d:\\diseaseFree.csv", row.names = F)
```