ISSN 2706-6819 Vol.6, Issue 6: 53-59, DOI: 10.25236/IJFM.2024.060608

Multifactor Cox Regression Model Analysis of Parotid Cancer Based on the SEER

Lixin Wang[1,*], Binying Huang[2]

[1]Development of Dentistry (Doctor of Dental Medicine), University of Baguio, Baguio, 2600, Philippines
[2]Second Clinical Medical School, Guangdong Medical University, Dongguan, Guangdong, 523109, China
[*]Corresponding author: 3203420371@qq.com

Abstract: *In this study, we aimed to explore the multifactor Cox regression model and its influencing factors for parotid cancer using the publicly available data and research resources from the Surveillance, Epidemiology, and End Results (SEER) database established by the American Cancer Research Center. We conducted a retrospective analysis and observed statistical data from 1653 cases of parotid cancer patients. We utilized a multifactor Cox regression model to screen for risk factors, evaluated the model using the C-index, assessed the accuracy of the 3-year and 5-year survival models through ROC curve analysis, and predicted the 3-year and 5-year survival probabilities using calibration plots. The results were presented using column line graphs. The multifactor Cox regression model analyzed age, gender, race, T stage, and N stage as risk factors for parotid cancer. The data revealed that the older the age, the higher the likelihood of developing parotid cancer, with a significantly higher proportion observed in White males compared to Black and Asian individuals. ROC analysis yielded an AUC of 0.84 for 3-year survival and 0.842 for 5-year survival. Parotid cancer, regardless of its benign or malignant nature, does not exhibit significant age restrictions, but it is commonly found in middle-aged and elderly populations. Clinical recommendations include regular monitoring of symptoms in parotid cancer patients, assessing T, N, M staging, and patient prognosis, with surgery being the optimal treatment modality for parotid cancer.*

Keywords: *Parotid Cancer; Survival Rate; Regression Analysis; Influencing Factors*

1. Introduction

The parotid gland is located on the lower side of the face, near the angle of the jaw[1], and is susceptible to the development of tumors from abnormal cell proliferation. These tumors can be classified as benign or malignant and are commonly identified clinically using fine needle aspiration cytology and frozen section techniques[2]. Parotid adenocarcinoma is a specific form of malignant tumor of the salivary glands that arises from the glandular tissues of the head and neck and has several subtypes, of which the mixed type is the most common[3][4]. This malignancy is relatively rare, accounting for only 0.5% of all cancers and less than 5% of head and neck cancers, with an incidence of only 0.4 to 1.2 cases per 100,000 people per year[5].

The course of parotid gland is usually long, and each of its subtypes exhibits a different biological behavior[6]. Clinically, patients may present with painless swelling of the parotid region, facial asymmetry due to nerve involvement, and possible ulceration around the tumor[7]. Surgical intervention is the cornerstone of the treatment of parotid cancer, and the treatment strategy should be based on the patient's age, clinical features of the tumor, size, degree of differentiation, and pathological staging[8][9]. These factors are not only closely related to patient survival outcomes, but also have a wide-ranging impact on the prognosis of patients with parotid adenocarcinoma[10]. Early detection and precise treatment can greatly improve the cure rate and survival of patients with parotid gland adenocarcinoma[11].

This study conducted a statistical analysis of 1653 cases of parotid cancer patients from the Surveillance, Epidemiology, and End Results (SEER) database, aiming to comprehensively explore their survival rates and associated risk factors.

ISSN 2706-6819 Vol.6, Issue 6: 53-59, DOI: 10.25236/IJFM.2024.060608

2. Materials and Methods

2.1 General Information

The data for this study were extracted from the SEER database, comprising a total of 1653 cases of parotid cancer patients. The patients' ages ranged from 1 to 85 years old, with 945 cases (57.1%) being male and 708 cases (42.8%) being female, indicating a slightly higher prevalence in males. Regarding ethnicity, the analysis revealed 1351 cases (81.7%) among White individuals, which significantly outnumbered cases among Black (7.0%) and Asian (11.2%) individuals.

2.2 Multifactor Cox Regression Model

The Cox regression model, also known as the proportional hazards regression model[12], is a statistical method used to analyze survival data, particularly when investigating the impact of various factors on survival time[12]. In survival analysis, the subjects under study typically experience an event (such as death or disease recurrence) within a certain period. The multifactor Cox regression model considers the influence of multiple predictor variables on the risk of events while controlling for other factors[13].

2.3 C-index

The C-index, also known as the Concordance index[14]， is a statistical metric used to assess the predictive ability of survival analysis models. It is employed to evaluate the accuracy and predictive power of predictive models. The C-index is a numerical value ranging between 0 and 1, where 1 indicates perfect prediction, 0.5 denotes random prediction, and values less than 0.5 suggest poor predictive performance[15]. It finds widespread application in medical research and clinical practice.

2.4 ROC Curve

The ROC (Receiver Operating Characteristic)[16] curve is a tool used to assess the performance of binary classification models, typically employed to measure the accuracy of model classification. The ROC curve plots the true positive rate against the false positive rate. In binary classification models, the true positive rate represents the proportion of actual positive cases that are correctly identified, while the false positive rate indicates the proportion of negative cases incorrectly classified as positive[17].

In the ROC curve, the Area Under the Curve (AUC) represents the area under the ROC curve, ranging from 0 to 1. A higher AUC value indicates better model performance, while an AUC of 0.5 suggests that the model's classification accuracy is equivalent to random guessing, indicating no predictive capability[18]. Therefore, AUC is a critical metric for evaluating the performance of binary classification models, providing a single value to quantify the overall classification accuracy of the model[19].

2.5 Nomogram Plot

The Nomogram plot is constructed based on multifactor regression analysis. It integrates multiple predictive indicators into scaled line segments, which are drawn on the same plane according to a certain proportion. All factors are scored, and the resulting scores are aggregated to predict the interrelationships between variables in the model[20].

3. Results

3.1 Multifactor Cox Regression Model Data Analysis

The data obtained from the multifactor Cox regression model using R-gui are presented as follows:

Table 1: Analysis of Cox Regression Model Results and Table 2: Factors with Confidence Intervals in Cox Regression Analysis

ISSN 2706-6819 Vol.6, Issue 6: 53-59, DOI: 10.25236/IJFM.2024.060608

Table 1: Analysis of Cox Regression Model Results

Category	coef	exp(coef)	se(coef)	z	Pr(>|z|)	
age50-59	9.354e-01	2.548e+00	2.116e-01	4.420	9.86e-06	***
age60-69	1.193e+00	3.298e+00	1.974e-01	6.043	1.51e-09	***
age70-79	1.873e+00	6.508e+00	1.893e-01	9.895	< 2e-16	***
age>=80	2.640e+00	1.402e+01	1.858e-01	14.214	< 2e-16	***
sexMale	2.957e-01	1.344e+00	8.341e-02	3.545	0.000393	***
raceWhite	4.537e-02	1.046e+00	1.740e-01	0.261	0.794295	
raceOther	7.702e-02	1.080e+00	2.116e-01	0.364	0.715844	
stage_TT1	1.349e+01	7.232e+05	8.940e+02	0.015	0.987960	
stage_TT2	1.404e+01	1.255e+06	8.940e+02	0.016	0.987468	
stage_TT3	1.440e+01	1.797e+06	8.940e+02	0.016	0.987148	
stage_TT4a	1.466e+01	2.318e+06	8.940e+02	0.016	0.986921	
stage_TT4b	1.481e+01	2.698e+06	8.940e+02	0.017	0.986785	
stage_TT4NOS	1.434e+01	1.682e+06	8.940e+02	0.016	0.987207	
stage_TTX	1.437e+01	1.736e+06	8.940e+02	0.016	0.987179	
stage_NN1	3.111e-01	1.365e+00	1.161e-01	2.680	0.007365	**
stage_NN2a	3.422e-01	1.408e+00	3.935e-01	0.870	0.384486	
stage_NN2b	5.042e-01	1.656e+00	9.997e-02	5.044	4.57e-07	***
stage_NN2c	8.515e-01	2.343e+00	3.605e-01	2.362	0.018190	*
stage_NN2NOS	2.016e+00	7.510e+00	7.201e-01	2.800	0.005113	**
stage_NN3	4.531e-01	1.573e+00	4.559e-01	0.994	0.320229	
stage_NNX	3.839e-01	1.468e+00	1.864e-01	2.060	0.039440	*
stage_MM1	1.187e+00	3.278e+00	1.348e-01	8.811	< 2e-16	***

Significance codes are utilized to signify the significance level of statistical outcomes:`***` signals a highly significant result with a p-value less than 0.001. `**` indicates a result is very significant, corresponding to a p-value less than 0.01. `*` denotes a significant result where the p-value is less than 0.05. `.` represents borderline significance with a p-value less than 0.1. `space` is used to denote a non-significant result, which corresponds to a p-value of 0.1 or greater.

Table 2: Factors with Confidence Intervals in Cox Regression Model Analysis

Category	exp(coef)	exp(-coef)	lower .95	upper .95
age50-59	2.548e+00	3.924e-01	1.6831	3.858
age60-69	3.298e+00	3.033e-01	2.2393	4.856
age70-79	6.508e+00	1.537e-01	4.4908	9.431
age>=80	1.402e+01	7.133e-02	9.7414	20.178
sexMale	1.344e+00	7.440e-01	1.1413	1.583
raceWhite	1.046e+00	9.556e-01	0.7440	1.472
raceOther	1.080e+00	9.259e-01	0.7134	1.635
stage_TT1	7.232e+05	1.383e-06	0.0000	Inf
stage_TT2	1.255e+06	7.967e-07	0.0000	Inf
stage_TT3	1.797e+06	5.566e-07	0.0000	Inf
stage_TT4a	2.318e+06	4.314e-07	0.0000	Inf
stage_TT4b	2.698e+06	3.706e-07	0.0000	Inf
stage_TT4NOS	1.682e+06	5.946e-07	0.0000	Inf
stage_TTX	1.736e+06	5.759e-07	0.0000	Inf
stage_NN1	1.365e+00	7.326e-01	1.0872	1.714
stage_NN2a	1.408e+00	7.102e-01	0.6512	3.045
stage_NN2b	1.656e+00	6.040e-01	1.3611	2.014
stage_NN2c	2.343e+00	4.268e-01	1.1559	4.750
stage_NN2NOS	7.510e+00	1.332e-01	1.8309	30.803
stage_NN3	1.573e+00	6.356e-01	0.6438	3.844
stage_NNX	1.468e+00	6.812e-01	1.0187	2.115
stage_MM1	3.278e+00	3.051e-01	2.5173	4.269

The concordance, a measure of predictive accuracy, is determined to be 0.805 with a standard error (SE) of 0.007. This suggests a relatively high level of predictive accuracy within the model.

Additionally, the likelihood ratio test yields a statistic of 859.9 on 22 degrees of freedom (df), with a

ISSN 2706-6819 Vol.6, Issue 6: 53-59, DOI: 10.25236/IJFM.2024.060608

p-value smaller than 2e-16, indicating strong evidence against the null hypothesis. Similarly, the Wald test produces a statistic of 756.1 on 22 df, with a p-value smaller than 2e-16, suggesting significant evidence against the null hypothesis. Moreover, the Score (logrank) test results in a statistic of 1032 on 22 df, with a p-value smaller than 2e-16, indicating substantial evidence against the null hypothesis.

Consequently, it can be inferred that higher hazard factor coefficients (exp(coef)) correlate with an increased risk of occurrence.

In the multifactor Cox model analysis, the concordance index remains at 0.805, with a standard error of 0.007, reaffirming the model's predictive accuracy.

3.2 Evaluation and Analysis of the Model Using C-index

Through C-index analysis, **(Table 3)** was derived, indicating a concordance index of 8.046576e-01. The results are similar to those obtained from the multifactor Cox regression model analysis. Moreover, the C-index results fall between 0.71 and 0.90, indicating moderate accuracy, suggesting that the model is error-free.

Table 3: C-index Analysis Table

C Index	Dxy	S.D.	n	missing	uncensored
8.046576e-01	1.609315e+00	9.851654e-01	-1.652000e+03	1.000000e+00	7.100000e+02
Relevant Pairs	Concordant	Uncertain			
-1.723629e+06	-3.366970e+05	-1.000343e+06			

3.3 Evaluation of the Accuracy of the 3-Year and 5-Year Survival Models Using ROC Curves

Figure 1 and Figure 2 respectively depict the ROC curves for the three-year and five-year survival rate models. From the figures, we can observe that the AUC for three-year survival is 0.84, while for five-year survival, it is 0.842. AUC values falling between 0.71 and 0.90 indicate moderate accuracy.

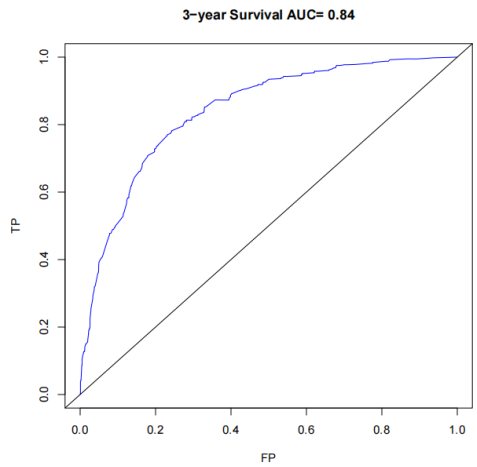

Figure 1: ROC Curve for the 3-Year Survival Model

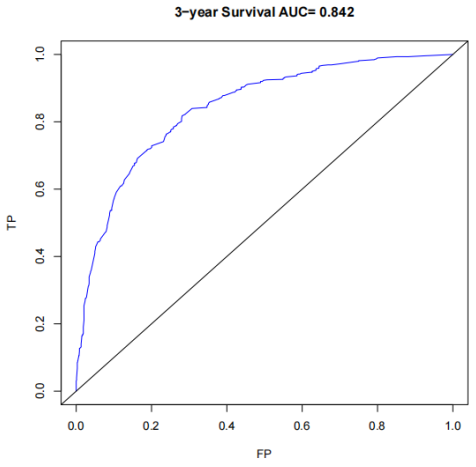

Figure 2: ROC Curve for the 5-Year Survival Model

Below are the forest plot (Figure 3: assessing the contribution of each influencing factor to the final variable by assigning scores to each factor, summing them up to obtain a total score, and predicting the individual final event probability through the function conversion rate between the total score and the final event probability) and survival curve analysis plots (Figures 4-7: risk score, age, race, gender) of the multifactor Cox regression analysis for salivary gland cancer.

ISSN 2706-6819 Vol.6, Issue 6: 53-59, DOI: 10.25236/IJFM.2024.060608

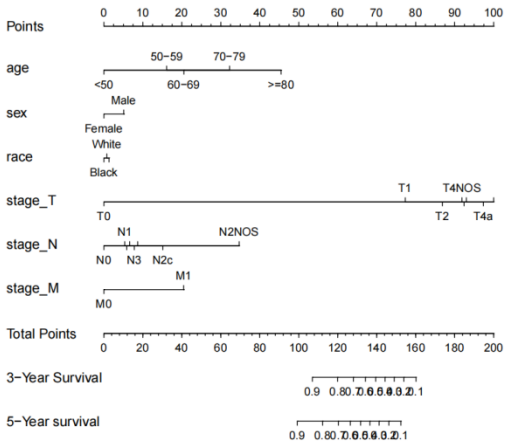

Figure 3: Forest Plot of Multifactor Cox Regression for Parotid Cancer

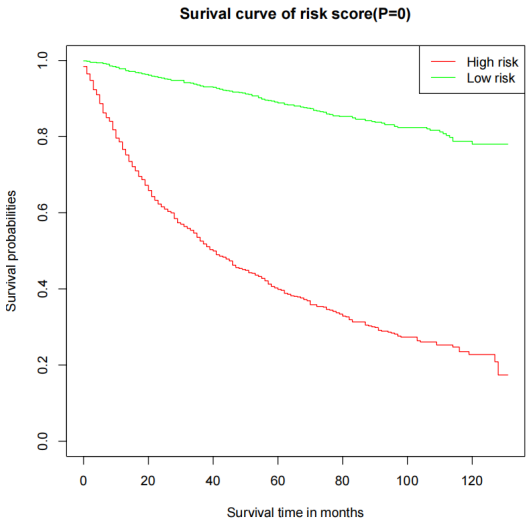

Figure 4: Risk Score Survival Curve

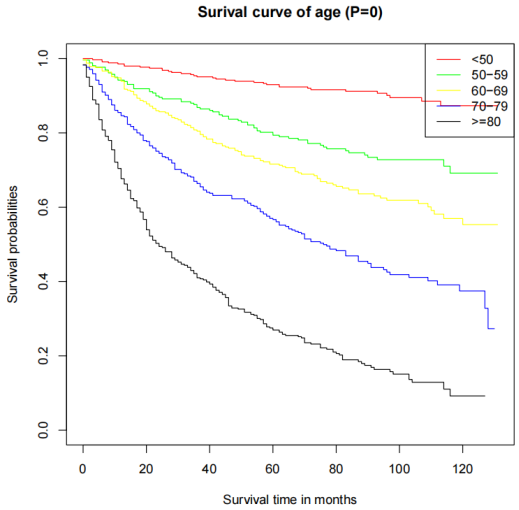

Figure 5: Age Survival Curve

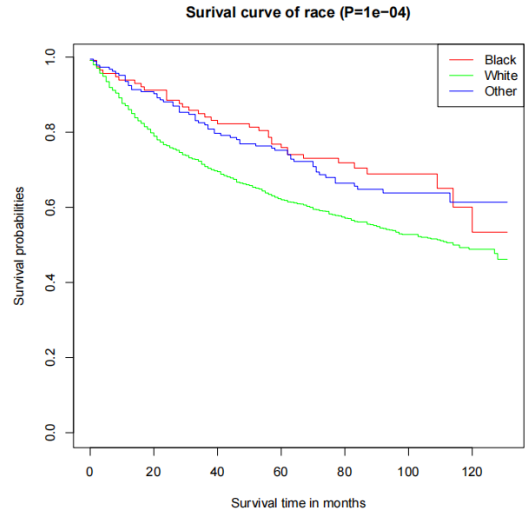

Figure 6: Race Survival Curve

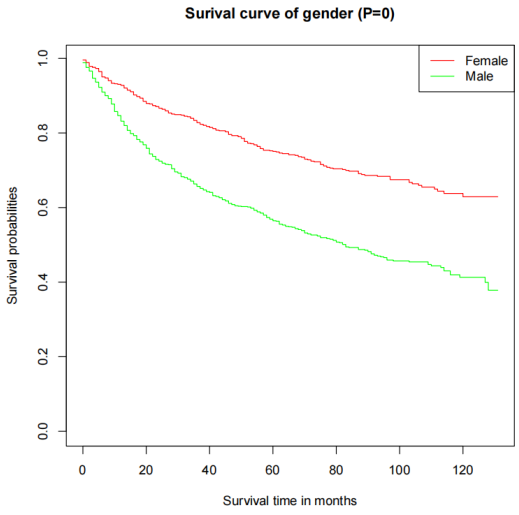

Figure 7: Gender Survival Curve

From Figure 4, it can be observed that the survival rate in the low-risk group is significantly higher compared to the high-risk group.

In Figure 5, concerning age, it is evident that the survival rate decreases with increasing age.

ISSN 2706-6819 Vol.6, Issue 6: 53-59, DOI: 10.25236/IJFM.2024.060608

Regarding race, as depicted in Figure 6, the survival rate of Caucasian individuals with cancer is notably lower compared to Black and Asian individuals.

Figure 7 indicates that males exhibit a significantly lower survival rate compared to females.

4. Conclusion

Through the analysis of the SEER database, we discovered that factors such as age, race, gender, tumor size, pathological type, differentiation degree, and lymph node metastasis all influence the risk factors and survival rates of patients with parotid cancer. Additionally, parotid cancer can be caused by factors such as genetics, environment, lifestyle habits, and endocrine factors. Multifactor Cox analysis revealed that tumor staging, age, and lymph node metastasis are the main factors affecting patient prognosis, with different impacts at different stages Early-stage parotid cancer often lacks obvious clinical symptoms and can only be detected during routine physical examinations. The preventive methods for parotid cancer are not yet clear and currently rely on maintaining regular routines, undergoing regular check-ups, and enhancing immune function.

Therefore, conducting risk factor analysis for parotid cancer can better equip doctors with the ability to predict the prognosis of different clinical patients. Furthermore, understanding how these factors influence the incidence, treatment outcomes, and survival rates of parotid cancer can contribute to prevention, diagnosis, and treatment strategies. Surgery is the optimal treatment for parotid cancer, and studies have shown that adjuvant radiotherapy after surgery can enhance efficacy. The extent of surgery is determined based on the size of the lesion, pathological type, and degree of malignancy. Early treatment and surgery significantly improve the survival rate of patients.

References

[1] Du Zhonghong. (2011). Clinical Treatment Analysis of 127 Cases of Parotid Gland Tumors. Journal of Oral Medicine Research, 27(5), 417-418.
[2] Huang Aiyu, Zhang Mengyin, & Xue Ming. (1991). Comprehensive Treatment of Advanced Parotid Cancer (with an Analysis of 41 Cases). Shanghai Medicine, 14(10), 575-579.
[3] Ye J, Li J. [Applied anatomic and biomechanical study on reconstruction of posterolateral complex of knee] [J]. Chinese journal of reparative and reconstructive surgery, 2010, 24(10):1199-1203. DOI:http://dx.doi.org/.
[4] Katz, A. D. (1975). Unusual lesions of the parotid gland. Journal of Surgical Oncology, 7(3), 219-235.
[5] Chang, J., Hong, H., Ban, M., Shin, Y., Kim, W., Koh, Y., & Choi, E. (2015). Prognostic Factors and Treatment Outcomes of Parotid Gland Cancer. Otolaryngology-Head and Neck Surgery, 153(6), 981-989.
[6] Shang Biao, Wei Qichun, Wang Kejing, Guo Liang, & Ge Minghua. (2008). Clinical Analysis of 135 Cases of Parotid Cancer. Zhejiang Medicine, 30(10), 1110-1111.
[7] Fan Dexin. (2004). Clinical Analysis of 168 Cases of Parotid Gland Tumors. Journal of Oral and Maxillofacial Surgery, 14(4), 354-356.
[8] Guo Liang, Wang Kejing, Liu Aihua, & Shang Jinbiao. (1999). Treatment Evaluation of 244 Cases of Parotid Cancer. Otolaryngology: Head and Neck Surgery, 6(3), 150-153.
[9] Ferreira, P. C., Amarante, J. M., Rodrigues, J. M., Pinho, C. J., Cardoso, M. A., & Reis, J. C. (2005). Parotid Surgery: Review of 107 Tumors. Int Surg, 200590, 160-166.
[10] Fan Fengyun, Wei Wensheng, Guo Yan, Shi Mei, & Xu Demen. (2004). Evaluation of Postoperative Radiotherapy for Parotid Cancer. Journal of Practical Oral Medicine, 20(6), 678-680.
[11] Choi, S. Y., Lee, E., Kim, E., Chung, M. K., Son, Y. I., Baek, C. H., & Jeong, H. S. (2021). Clinical outcomes of bulky parotid gland cancers: need for self-examination and screening program for early diagnosis of parotid tumors. BMC cancer, 21, 1-9.
[12] Liu, Z., Wang, Q., & Zhao, Y. (2019). Expression and Clinical Significance of P16, Ki-67, and CyclinD1 in Oral Tissues of Oral Cancer Patients. Modern Medicine, 47, 1-4.
[13] Schlichting, P., Christensen, E., Andersen, P. K., Fauerholdt, L., Juhl, E., Poulsen, H., ... & Copenhagen Study Group for Liver Diseases. (1983). Prognostic factors in cirrhosis identified by Cox's regression model. Hepatology, 3(6), 889-895.
[14] Longato, E., Vettoretti, M., & Di Camillo, B. (2020). A practical perspective on the concordance index for the evaluation and selection of prognostic time-to-event models. Journal of biomedical informatics, 108, 103496.

ISSN 2706-6819 Vol.6, Issue 6: 53-59, DOI: 10.25236/IJFM.2024.060608

[15] Hartman, N., Kim, S., He, K., & Kalbfleisch, J. D. (2023). Pitfalls of the concordance index for survival outcomes. Statistics in Medicine, 42(13), 2179-2190.
[16] Hoo, Z. H., Candlish, J., & Teare, D. (2017). What is an ROC curve? Emergency Medicine Journal, 34(6), 357-359.
[17] Metz, C. E. (1978, October). Basic principles of ROC analysis. In Seminars in nuclear medicine (Vol. 8, No. 4, pp. 283-298). WB Saunders.
[18] Narkhede, S. (2018). Understanding auc-roc curve. Towards data science, 26(1), 220-227.
[19] Marzban, C. (2004). The ROC curve and the area under it as performance measures. Weather and Forecasting, 19(6), 1106-1114.
[20] Zhang, Z., & Kattan, M. W. (2017). Drawing Nomograms with R: applications to categorical outcome and survival data. Annals of translational medicine, 5(10).