Biomolecules and Biomedicine

ISSN: 2831-0896 (Print) | ISSN: 2831-090X (Online)
Journal Impact Factor® (2024): 2.2

CiteScore® (2024): 5.2
www.biomolbiomed.com | blog.bjbms.org

RESEARCH ARTICLE

Du et al: Gastrointestinal DLBCL: Clinical and prognostic analysis

Gastrointestinal diffuse large B-cell lymphoma: Clinical characteristics and prognostic analysis from SEER database

Fang Du[1#], Lingyun Zhou[1#], Runya Fang[2], Jiao Chen[1], Danbo Liu[1], Hongxian Xiang[1], Wenyi Lu[1], Jingsong Wu[1], Haifei Chen[1*]

[1]Department of Hematology, Luohu People's Hospital of Shenzhen, Shenzhen, China

[2] Guangzhou Institute of Cancer Research, the Affiliated Cancer Hospital, Guangzhou Medical University, Guangzhou, China

*****Correspondence to Haifei Chen: Chhf1224@163.com

[#]Equally contributed to this work: Fang Du and Lingyun Zhou.

DOI: https://doi.org/10.17305/bb.2025.12697

ABSTRACT

This study systematically analyzed the clinicopathological characteristics and prognostic factors of gastrointestinal diffuse large B-cell lymphoma (GI-DLBCL) patients using the SEER database. The Kaplan-Meier method was used to survival analysis, while LASSO regression analysis was utilized to further filter variables. The Pi for interaction was applied to verify the interactions in the multivariate analysis, and total survival risks were distinguished using hierarchical survival curves. Multivariate Cox regression analysis revealed that hazard ratio (HR) values indicated that age over 60 years (HR = 2.85), Ann Arbor stage (stage II: HR = 1.22; stage III: HR = 1.31; stage IV: HR = 1.85), and being widowed (HR = 1.40) were independent poor prognostic factors. In contrast, chemotherapy (HR = 0.37), radiotherapy (HR = 0.84), surgery (HR = 0.86), and lymph node resection (HR = 0.79) were associated with significant survival benefits. Additionally, an intestinal primary site (HR = 0.89), white race (HR = 0.78), and other races (HR = 0.65) were correlated with better prognosis. The nomogram model constructed from these independent prognostic factors demonstrated excellent predictive performance in both the training and validation cohorts, achieving a C-index of 0.71, significantly outperforming the traditional Ann Arbor staging system, which had a C-index of 0.56. Receiver operating characteristic (ROC) curve analysis indicated high discriminative ability for predicting 3-year, 5-year, and 10-year survival rates, with area under curve (AUC) values of 0.746, 0.756, and 0.756, respectively. Decision curve analysis (DCA) further confirmed the model's significant clinical net benefit across a wide range of threshold probabilities. The nomogram model developed in this study, based on extensive SEER database data, effectively predicts the prognosis of GI-DLBCL patients and provides a quantitative tool for individualized treatment.

KEYWORDS: Gastrointestinal diffuse large B-cell lymphoma; SEER database; Prognostic factors; Nomogram; Survival analysis.

INTRODUCTION

The GI-DLBCL is one of the most common types of extranodal non Hodgkin lymphoma, accounting for approximately 30% -40% of all extranodal lymphomas [1,2]. The stomach and intestines are mainly primary sites of GI-DLBCL, which have significantly heterogeneous clinical manifestations and prognosis. Some patients are still prone to recurrence orprogression despite receiving standard chemotherapy, with a 5-year overall survival rate of less than 60% [3,4]. The prognosis of patients with refractory diffuse large B-cell lymphoma is even worse, with a median overall survival of only 6.3 months and a 2-year survival rate of 20% [5]. Currently, the Ann Arbor staging System is the cornerstone for lymphoma staging, but its prognostic predictive performance for extranodal lesions, especially GI-DLBCL, remains controversial [6]. The Ann Arbor system may underestimate the impact of primary site, local treatment (such as surgery or radiotherapy), and patient baseline characteristics (such as age and comorbidities) on prognosis [7], which result in challenges in formulating individualized treatment strategies in clinical practice. Previous study suggest that the Ann Arbor system may be less effective than the TNM staging system in predicting overall survival for patients with primary gastrointestinal lymphoma [8]. However, there is still a lack of effective prognostic prediction models for GI-DLBCL patients.

The nomogram-based approaches and SEER-based prognostic models for DLBCL have been been widely reported [9-14]. Wang et.al. studied prognostic models of GI DLBCL based on SEER database, Kaplan-Meier survival curves and nomogram, they mainly focused on the analysis of overall survival (OS) and median OS [9]. Liu et.al. studied SEER-based prognostic models for primary small intestinal DLBCL, which indicate chemotherapy and surgery are beneficial to survival [10]. Primary gastric DLBCL based on SEER database from 1973 to 2014, multivariable analysis revealed that the cases in 2001-2014 has lower mortality (HR=0.892, p=0 001) [11]. Wang et.al. establish an SEER-based prognostic model by dynamic prognostic nomogram to predict the OS of elderly patients with GI DLBCL [12]. The research of small intestine and colon DLBCL based on SEER database of 1613 cases indicate that age, Ann Arbor stage, divorced or separated, uninsured and primary colon were associated with prognosis [13]. Feng et.al. studied SEER-based primary GI-DLBCL by using Kaplan-Meier curves and Cox regression analysis, which indicate that five-year OS rates of stomach, small intestine, and colorectum are about 50%, with corresponding cancer-speciffc survival rates of about 65% [14]. Multivariate Cox regression suggest age, race, marital status, tumor stage, location, and treatment as independent risk factors.

In recent years, prognostic model studies based on large-sample databases have provided new insights for precision medicine in oncology. The SEER database is an important tool for exploring prognostic factors in rare cancers due to its broad population coverage, long-term follow-up, and detailed clinical variables[15-17]. This study aims to integrate clinical data of GI-DLBCL patients from the SEER database between 2004 and 2020, analyze their clinical characteristics and the impact on prognosis. Based on analyzed data, a quantifiable nomogram predictive model to address the lack of effective prognostic assessment tools for GI-DLBCL was constructed, which providing evidence-based guidance for identifying high-risk patients and optimizing follow-up strategies, and further advancing precision medicine in this field.

MATERIALS AND METHODS

Data sources and research population

The data for this study were sourced from the SEER database established by the US National Cancer Institute. Patients diagnosed with GI-DLBCL between 2004 and 2020 were retrieved by SEER*Stat software (version 8.4.1). The inclusion criteria included: (1) confirmed pathological type as DLBCL; (2) primary site as stomach or intestine; (3) complete clinical information and follow-up data. The exclusion criteria included: (1) unknown race, lymph node dissection status, surgical and marital status; (2) unknown Ann Arbor stage; (3) unknown survival follow-up time. The screening process is shown in Figure 1.

Data collection

The basic demographic data and clinicopathological features of patients were collected, which including: age, sex, race (white, black, other), marital status, primary site (stomach or intestine), Ann Arbor stage, whether chemotherapy, radiotherapy, surgery, lymph node resection (LNR), income level (≤ 75,000 USD or > 75,000 USD). The predefined research end point is overall survival (OS), that is, the time from diagnosis to death or the last follow-up. Survival state and survival time are provided by SEER database.

Statistical analysis

Statistical analysis was carried out in R language (version 4.2.2). Counting data were expressed by frequency and percentage, and χ^2 test was used for comparison between groups. Kaplan-Meier method was used to draw the survival curve, and Log-rank test was used to compare the survival differences among different variables. LASSO regression analysis was

used to further filter variables, Pi for interaction was used to verify the interaction of multivariate, total survival risks was distinguished by hierarchical survival curves. Furthermore, Cox proportional hazard regression model was used for univariate and multivariate analysis to select independent risk factors affecting overall survival. According to the results of multivariate Cox regression, nomogram were constructed to predict the survival probability of individuals in 3 years, 5 years and 10 years. The discriminant ability of the model is evaluated by consistency index (C-index) and the area under curve (AUC) of receiver operating characteristic curve (ROC). Then the consistency between the predicted survival probability of the model and the actual observed survival probability is further verified by calibration curve. At the same time, the decision curve analysis (DCA) is used to quantify the clinical net benefit of the model under different risk thresholds, so as to evaluate its clinical application value. The differences are considered statistically significant when $P < 0.05$.

RESULTS

Basic clinical features

A total of 4437 patients with gastrointestinal diffuse large B-cell lymphoma (GI-DLBCL) diagnosed in SEER database from 2004 to 2020 were randomly divided into training group (3105 cases) and verification group (1332 cases) according to the ratio of 7:3. In all the patients, 59.91% patients were male and 40.09% were female. 71.87% patients are over 60 years old. Race is mostly white (80.32%), while other races and blacks account for 13.16% and 6.51% respectively. In terms of primary site, stomach and intestine accounted for 50.98% and 49.02% respectively. Ann Arbor stages are mainly I (43.79%) and II (26.77%) stages. 69.53% patients received chemotherapy, 36.74% patients underwent surgery and 17.80% patients underwent lymphadenectomy. 40.43% patients had an income exceeding $75,000. The training group and the verification group are evenly distributed in the above variables, with no statistically significant differences (all $P \geq 0.05$), as shown in Table 1.

Influencing factors of prognosis in patients with GI-DLBCL

Univariate Cox regression analysis showed that age, race, radiotherapy, chemotherapy, operation, LNR, primary site, Ann Arbor stage and marital status were significantly related to the prognosis of patients with GI-DLBCL ($P < 0.05$). Multivariate analysis further screened out the following independent prognostic factors: age > 60 years old (HR=2.85, 95% CI: 2.49 ~ 3.26, $P < 0.001$), and late Ann Arbor stage (HR=1.31 in stage III, HR=1.85 in stage IV, P <

0.001) have a poor prognosis. For age > 60 years old patients (HR=2.85, 95% CI: 2.49-3.26, P<0.001) have a mortality risk (HR) 2.85 times higher than other patients. After filtering for age, widowed was still significantly associated with mortality risk. In addition, in the interactive testing, the Pi for interaction between marital status and age was 0.099 (see in supplementary materials), indicating that the interaction effect was not significant. Therefore, widowed (HR=1.40, 95% CI: 1.14~1.72, P=0.001) is an independent prognostic factor. The true hazard ratio (HR) has a 95% probability of being between the confidence interval (CI) 2.49 and 3.26 times. Since CI does not include 1 and P<0.001, it indicates that age has a highly statistically significant impact on prognosis, and the effect value significantly deviates from "no impact" (when HR=1). Patients who did underwent chemotherapy (HR=0.37, P<0.001), radiotherapy (HR=0.84, P=0.023), surgery (HR=0.86, P=0.032), LNR (HR=0.79, P=0.002), or had a primary site in the intestine (HR=0.89, P=0.037) had better prognoses. In addition, compared with blacks, the prognosis of whites and "other" races is better (HR is 0.65 and 0.78 respectively, P < 0.01), as shown in Table 2. For the race, gender, marital status, Ann Arbor stage II and III, and income>75000 cases, the HR range includes 1 and P ≥ 0.05 because the survival is influenced by other multiple factors (such as age and treatment method), the short follow-up time, and the small samples (such as separated, black, Ann Arbor stage III), resulting in the impact on the survival being within the range of risk and protective factors. The high income level (>$75,000, HR=0.91) indicate that it is a protective factor in the univariate analysis, but the P=0.056 suggest it not statistically significant, thus the multivariate analysis is not been studied.

When constructing a Cox proportional hazards regression model, variable selection should balance statistical significance ($P < 0.05$) with clinical and biological rationality to avoid false associations or overfitting caused by simple stepwise regression. Lasso regression compresses variable coefficients β through L1 regularization (Lasso penalty term). As the penalty intensity λ increases, some variable coefficients are compressed to 0, achieving variable selection. As shown in Figure 2, Lasso regression uses L1 regularization to select variables, and ultimately selects Age (β=0.774, positively correlated with risk factor) and Chemotherapy (β=-0.733, negatively correlated with risk factor) as the two most significant variables, which was consistent with multivariate Cox regression analysis results of age >60 years (HR=2.85) and chemotherapy (HR=0.37). Ann Arbor Stage (β=0.235), Primary site (β=0.146), and Marital (β=0.157) have a positive impact on risk factors. Surgery (β=-0.302) and LNR (β=-0.114) may improve prognosis. While the absolute values of radiotherapy (β=-

0.02) and income (β=-0.04) are close to 0, and the β values of sex and race are 0, indicating a very little impact on risk factors. Compared with the Cox regression results in Table 2, the Ann Arbor Stage, marital, race, and radiotherapy in Lasso regression analysis have inconsistent effects on risk factors because Lasso regression did not consider variable stratification. For example, marital (β=0.157) shows a positive impact on the risk factor in Lasso regression, while in Cox regression, different marital statuses have different effects on the risk factor, include married (P=0.56), separated (P=0.052), single (P=0.152), widowed (P<0.001), indicating that only widowed has a negative impact on survival rate.

The survival curves of the above 9 independent prognostic factors are shown in Figure 2 and Figure 3. Among them, the survival rate of the elderly, late Ann Arbor stage and widowed patients showed a significant downward trend. The multivariate survival situation are analyzed by the Kaplan Meier method, which include 9 independent prognostic factors (P<0.05): race, marital status, chemotherapy, age, radiotherapy therapy, LNR, primary site, surgery, Ann Arbor stage. The survival differences between different groups are statistically significant, indicating that these factors can independently predict patient prognosis.

The Age, income, marital status, and race may be associated with competition risk, and univariate Cox models may not be sufficient to attribute the effects of these variables. The Fine-Gray sub distribution and multivariate Cox model risk analysis indicate that age and treatment method, race/income and treatment accessibility, LNR and Ann Arbor stage, marital status and treatment compliance have competitive risks for survival. For example, the proportion of widowed patients over 60 years old is higher, elderly patients ($>$ 60 years) have poor tolerance to radiotherapy and chemotherapy, low-income or ethnic minority patients may interrupt treatment due to economic limitations and face higher burden of complications, married patients have higher treatment compliance, and late Ann Arbor stage lesions are more prone to systemic metastasis. The interaction P_i value (P_i for interaction) is used to determine whether a variable significantly affects the effect of another variable, which is assessed by the statistical differences of HR between different subgroups. $P_i<0.05$ represents significant interaction, and $P_i \geq 0.05$ represents no significant interaction. In the interaction analysis between radiotherapy and other variables, only primary site ($P_i<0.05$) is interacted with radiotherapy. In the interaction analysis between chemotherapy and other variables, both the Ann Arbor stage and primary site with Pi<0.05 is interacted with chemotherapy. In the

interaction analysis between surgery and other variables, both the primary site and LNR with $P_i<0.05$ is interacted with surgery. In the interaction analysis between LNR and other variables, the Pi values of primary site, Ann Arbor stage, and surgery are less than 0.05, indicating significant interaction. 5. In the interaction test analysis between age and other variables, only the interaction test Pi values of Sex and Ann Arbor stage were less than 0.05, indicating significant interaction.

The multivariate factor related survival curve analysis (Figure 5) shows that the survival curve trends of the training and validation groups are highly consistent, indicating that the "risk stratification" effect can be repeated in different groups. The survival probability exhibits a gradient decrease as time for four groups (low, low media, high media, high) in both Figure 5 A and B. The inter group differences are highly statistically significant due to $p<0.0001$. Stratification based on multivariate comprehensive analysis can effectively distinguish the survival risk of patients. The low group is the low-risk group with high survival probability and slow event accumulation, while the high group is the high-risk group with low survival probability and fast event accumulation.

Establishment and verification of nomogram

According to the independent prognostic factors selected by multivariate Cox regression analysis, including age, race, primary site, surgery, LNR, radiotherapy, chemotherapy, marital status and Ann Arbor stage, nomogram model was constructed to predict the 3-year, 5-year and 10-year overall survival rates of patients with gastrointestinal DLBCL, as shown in Figure 6. Each variable is given a corresponding score according to its regression coefficient, and the individual survival probability can be predicted by adding up the scores of patients. In the nomogram, the points of 0-100 represent the univariate disk of death, and the total points of 0-450 represent the superposition of multivariate risks of death. The current axes ordering allocate the higher points for the higher-risk category. For example, whites and "other" races have better the prognosis than blacks, so the blacks have highest disk of death that result in highest points in different races.

In terms of model performance evaluation, the nomogram constructed in this study achieved a C-index of 0.71 (95% CI: 0.70-0.73) in the training group and similarly 0.71 (95% CI: 0.69-0.73) in the validation group, which indicate that this model has excellent discrimination capability. Herein, the difference of Nomogram and Ann Arbor stage is defined as ΔC, and the calculated ΔC is 0.15. Compared to traditional Ann Arbor stage, the predictive

performance of nomograms exhibit a relative improvement of approximately 26.8%. In contrast, the traditional Ann Arbor staging System had a C-index of only 0.56 (95% CI: 0.54-0.57), with a statistically significant difference ($P < 0.001$) (Table 3).

The ROC curve also shows that the nomogram constructed in this study has a good discrimination ability in predicting the overall survival, which is better than the traditional Ann Arbor stage model. Figure 7 compares the predictive performance of the column chart model with the Ann Arbor staging System in the training queue (A, C) and internal validation queue (B, D), which confirm that the column chart model exhibits significant advantages at different time points (3/5/10 years). Significantly, in long-term prognosis assessment (10-year survival), the column chart still maintains a stable performance with AUC > 0.73, while the predictive performance of traditional staging systems significantly declines over time (10-year AUC is only 0.56). The 95% confidence intervals for all AUC values are shown in Table 4.

Figure 8 shows the calibration curve analysis of the nomogram model, the dashed line represents the perfect prediction state, and the closer the scatter distribution is to the dashed line, the higher the consistency between the predicted probability of the model and the actual survival rate. For the training queue (A-C), the predicted lines for 3, 5, and 10 years all closely follow the dashed line, indicating that the model has excellent fitting performance in the training data. For the validation queue (D-F), the overall trend of the predicted line is consistent with the dashed line, but the deviations are larger than the training queue. The calibration curve shows that the total survival probability predicted by the nomogram model is highly consistent with the actual observation value, and has a good calibration degree.

Figure 9 shows the decision curve analysis (DCA) of of the nomogram model, the "all" reference strategy means treatment in all cases, while the "none" reference strategy means no treatment. The clinical applicability of the GI-DLBCL patient survival prediction column chart model in the training and internal validation cohorts is verified by comparing the net benefits of three decision strategies at different risk thresholds. The DCA results show that this model has a positive net benefit, suggesting that the model constructed in this study has good clinical application potential.

DISCUSSION

Based on the large sample data of SEER database, this study systematically analyzed the clinical characteristics and prognostic factors of patients with GI-DLBCL, and constructed a

nomogram model integrating variables such as age, stage, treatment method and social demography. The results showed that age > 60 years, late Ann Arbor stage and widowed status were independent adverse prognostic factors, while receiving systematic treatment (chemotherapy, radiotherapy, surgery and lymphadenectomy) and primary intestinal tract were significantly related to the improvement of survival.

Firstly, this study demonstrates that advanced age (>60 years) has a significant negative impact on prognosis (HR=2.85). Gao et.al. demonstrated that advanced age was also identified as a critical factor influencing prognosis [18]. This may be related to the decline in immune function, increased comorbidities, and reduced treatment tolerance in elderly patients. In addition, elderly patients with DLBCL are more likely to suffer from the decline of curative effect due to insufficient dose adjustment or interruption of treatment. Wang et al. developed a more accurate and convenient dynamic prognostic nomogram model specifically for elderly GI-DLBCL patients [12]. Secondly, the Ann Arbor staging System, as a traditional prognostic indicator, still shows independent predictive value in this study. However, the mortality risk (HR=1.85) for stage IV patients is significantly higher than previously reported [19], suggesting that the extranodal aggressiveness of GI-DLBCL may be underestimated. Further stratification using imaging or molecular markers is required [7]. Notably, widowed status (HR=1.40) was identified for the first time as an independent risk factor, potentially associated with insufficient social support, psychological stress, or delayed medical care [20, 21]. Further studies are needed to validate its biological or social mechanisms. Additionally, both cross-sectional and longitudinal studies indicate a significant association between widowhood and cognitive function decline, such as impaired executive function and reduced decision-making ability [22]. This may further weaken patients' adaptability to complex treatment regimens, increase disease management burdens, and affect the outcome of the disease.

At the therapeutic level, the survival benefit of chemotherapy is the most significant (HR=0.37), which is consistent with the core position of R-CHOP regimen in DLBCL [23]. However, the HR value of radiotherapy and surgery is close to 1 (0.84 and 0.86), suggesting that its effect may be limited to a specific subgroup (such as patients with locally advanced stage or chemotherapy intolerance) [24]. The prognostic value of lymphadenectomy (HR=0.79) provides a new basis for clinical practice, which may be related to reducing tumor load or improving local control. However, it is necessary to be wary of selection bias, such as

better baseline status of surgical patients. In addition, the better prognosis of primary intestinal patients (HR=0.89) may be related to the low concealment of early symptoms and the high timeliness of diagnosis, but the potential impact of anatomical differences (such as blood supply and microenvironment) still needs to be explored.

The nomogram model constructed in this study is superior to the Ann Arbor staging System in both discrimination (C-index=0.71) and calibration (C-index=0.56), with decision curves confirming significant clinical net benefit. This result is consistent with the trend of extensive application of nomogram in solid tumors [9, 25]. The advantages of the nomogram model lie in incorporating treatment methods and social factors, which provides a quantitative tool for individualized prognosis evaluation. However, this model still has limitations: first, SEER database lacks key variables such as ECOG score, LDH level and molecular typing (such as GCB/ABC subtype), which may affect the comprehensiveness of prediction; Second, retrospective design can't avoid the interference of confounding factors (such as biases in treatment choice); Thirdly, the external verification of the model needs to include multi-center or different ethnic groups to improve universality.

CONCLUSION

In summary, , through the analysis of large sample data, this study identified the independent effects of age, Ann Arbor stage, treatment methods, and social factors on the prognosis of gastrointestinal DLBCL, and successfully constructed the nomogram prediction model for this disease.The collected clinical variables include demographic characteristics, disease characteristics, and treatment information, with more stratified analysis used for marital status, Ann Arbor staging, race, and treatment methods. Multivariate Cox regression analysis suggested that age>60 years old (HR=2.85), Ann Arbor stage (stage II: HR=1.22; Phase III: HR=1.31; Stage IV: HR=1.85) and widowed status (HR=1.40) are independent poor prognostic factors. Different treatment methods include chemotherapy (HR=0.37), radiotherapy (HR=0.84), surgery (HR=0.86), and lymph node dissection (HR=0.79), exhibited significant survival benefits. Intestinal primary (HR=0.89), white race(HR=0.78) andother races (HR=0.65) were also associated with better prognosis. Lasso regression analysis selected age (β=0.774, positively correlated with risk factors) and chemotherapy (β=-0.733, negatively correlated with risk factors) as the two most significant variables, which was consistent with the multivariate Cox regression analysisresults. Through the analysis of the interaction between radiotherapy, chemotherapy, surgery, LNR, age, and other variables,

which is shown that the primary site and Ann Arbor staging are prone to interactive effects with other variables. The nomogram model in both training and validation cohorts exhibited excellent predictive performance with a C-index of 0.71, that is significantly better than traditional Ann Arbor staging system (C-index=0.56). Moreover, stratified analysis of multivariate survival curves can effectively distinguish survival risks of patients. Therefore, this model providing support for identifying high-risk patients, adjusting treatment intensity, and optimizing follow-up strategies, and can quantitatively evaluate the survival probability of individual patients.. Future researches should focus on improving models through prospective studies and integrating molecular biomarkers to promote the application of precision medicine in the field of gastrointestinal lymphoma.

ACKNOWLEDGMENTS

The authors thank all the staff at the National Cancer Institute(USA) for their contributions to the SEER database.

Conflicts of interest: Authors declare no conflicts of interest.

Funding: This study was funded by Sanming Project of Medicine in Shenzhen (SZSM202301035), Guangdong Yiyang Healthcare Charity Foundation (JZ2024026).

Data availability: The original data is obtained from SEER database. All analysis data in this study are available from the corresponding or first authors on reasonable request.

Submitted: 23 May 2025

Accepted: 1 July 2025

Published online: 8 July 2025

REFERENCES

[1]Swerdlow S H, Campo E, Pileri S A, et al. The 2016 revision of the World Health Organization classification of lymphoid neoplasms [J]. Blood, 2016, 127(20): 2375-2390.https://doi.org/ 10.1182/blood-2016- 01-643569.

[2]Sharma B, Pavelock N, Antoine M, et al. Primary Diffuse Large B-Cell Lymphoma of the Descending Colon [J]. Am J Med Sci, 2019, 358(2): 164-167.https://doi..org/10.1016/j.amjms.2019.05.004

[3]Baliko A, Szakacs Z, Kajtar B, et al. Clinicopathological analysis of diffuse large B-cell lymphoma using molecular biomarkers: a retrospective analysis from 7 Hungarian centers [J]. Front Oncol, 2023, 13: 1224733.https://doi. org/10.3389/fonc.2023.1224733

[4]Chu Y, Liu Y, Jiang Y, et al. Prognosis and complications of patients with primary gastrointestinal diffuse large B-cell lymphoma: Development and validation of the systemic inflammation response index-covered score [J]. Cancer Med, 2023, 12(8): 9570-9582. https://doi.org/10.1002/cam4.5733

[5]Crump M, Neelapu S S, Farooq U, et al. Outcomes in refractory diffuse large B-cell lymphoma: results from the international SCHOLAR-1 study [J]. Blood, 2017, 130(16): 1800-1808.https://doi.org/10.1182/blood-2017-03-769620

[6]Hwang H S, Yoon D H, Suh C, et al. Intestinal diffuse large B-cell lymphoma: an evaluation of different staging systems [J]. J Korean Med Sci, 2014, 29(1): 53-60.https://doi.org/10.3346/jkms.2014.29.1.53

[7]McCarten K M, Nadel H R, Shulkin B L, et al. Imaging for diagnosis, staging and response assessment of Hodgkin lymphoma and non-Hodgkin lymphoma [J]. Pediatr Radiol, 2019, 49(11): 1545-1564.https://doi.org/10.1007/s00247-019-04529-8

[8]Chang S, Shi X, Xu Z, et al. TNM staging system may be superior to Lugano and Ann Arbor systems in predicting the overall survival of patients with primary gastrointestinal lymphoma [J]. J BUON, 2015, 20(3): 812-819.

[9]Wang J, Zhou M, Zhou R, et al. Nomogram for Predicting the Overall Survival of Adult Patients With Primary Gastrointestinal Diffuse Large B Cell Lymphoma: A SEER- Based Study [J]. Front Oncol, 2020, 10: 1093.https://doi.org/10.1038/s41598-021-02597-1

[10]Liu X, Cao D, Liu H, et al. Clinical Features Analysis and Survival Nomogram of Primary Small Intestinal Diffuse Large B-Cell Lymphoma [J].Cancer Manag Res, 2022,14: 2639-2648.https://doi.org/10.2147/CMAR.S369086

[11]Liu P, Xia Y, Bi X, et al. Trends in Survival of Patients with Primary Gastric Diffuse Large B-Cell Lymphoma: An Analysis of 7051 Cases in the SEER Database [J]. Dis Markers, 2018, 7473935. https://doi.org/10.1155/2018/7473935

[12]Wang J, Ren W, Zhang C, et al. A New Staging System Based on the Dynamic Prognostic Nomogram for Elderly Patients With Primary Gastrointestinal Diffuse Large B-Cell Lymphoma [J]. Front Med (Lausanne), 2022, 9: 860993.https://doi.org/10.3389/fmed.2022.860993

[13]Wang Y , Song J, Wen S, Zhang X. A visual model for prognostic estimation in patients with primary diffuse large B-cell lymphoma of small intestine and colon: analysis of 1,613 cases from the SEER database [J].Transl Cancer Res 2021,10(4):1842-1855.http://dx.doi.org/10.21037/tcr-20-3086

[14]Feng Y, Zheng S, Sun Y, Liu L. Location-specific analysis of clinicopathological characteristics and long-term prognosis of primary gastrointestinal diffuse large B-cell lymphoma [J]. Sci Rep-UK 2025, 15:19574.https://doi.org/10.1038/s41598-025-04537-9

[15] Nie Y, Ying B, Lu Z, et al. Predicting survival and prognosis of postoperative breast cancer brain metastasis: a population-based retrospective analysis [J]. Chin Med J (Engl), 2023, 136(14): 1699-1707.https://doi.org/10.1097/CM9.0000000000002674

[16] Lamba N, Wen P Y, Aizer A A. Epidemiology of brain metastases and leptomeningeal disease [J]. Neuro Oncol, 2021, 23(9): 1447-1456.https://doi.org/10.1093/neuonc/noab101

[17]Chu T, Meng Y, Wu P, et al. The prognosis of patients with small cell carcinoma of the cervix: a retrospective study of the SEER database and a Chinese multicentre registry [J]. Lancet Oncol, 2023, 24(6): 701-708.https://doi. org/10.3389/fonc.2023.1224733

[18]Gao F, Wang Z F, Tian L, et al. A Prognostic Model of Gastrointestinal Diffuse Large B Cell Lymphoma [J]. Med Sci Monit, 2021, 27: e929898.https://doi.org/10.12659/MSM.929898

[19]Zhou Z, Sehn L H, Rademaker A W, et al. An enhanced International Prognostic Index (NCCN-IPI) for patients with diffuse large B-cell lymphoma treated in the rituximab era [J]. Blood, 2014, 123(6): 837-842.https://doi.org/10.1182/blood-2013-09-524108

[20]Rodin R, Smith A K, Espejo E, et al. Mortality and Function After Widowhood Among Older Adults With Dementia, Cancer, or Organ Failure [J]. JAMA Netw Open, 2024, 7(9): e2432979.https://doi.org/10.1001/jamanetworkopen.2024.32979

[21]Lagergren J, Andersson G, Talback M, et al. Marital status, education, and income in relation to the risk of esophageal and gastric cancer by histological type and site [J]. Cancer, 2016, 122(2): 207-212.https://doi.org/10.1002/cncr.29731

[22]Wu-Chung E L, Leal S L, Denny B T, et al. Spousal caregiving, widowhood, and cognition: A systematic review and a biopsychosocial framework for understanding the relationship between interpersonal losses and dementia risk in older adulthood [J]. Neurosci Biobehav Rev, 2022, 134: 104487.https://doi.org/10.1016/j.neubiorev.2021.12.010

[23] Wilson W H, Wright G W, Huang D W, et al. Effect of ibrutinib with R-CHOP chemotherapy in genetic subtypes of DLBCL [J]. Cancer Cell, 2021, 39(12): 1643-1653 e1643.https://doi.org/10.1016/j.ccell.2021.10.006

[24] Zhang C, Zhang X, Liu Z, et al. The impact of surgery on long-term survival of patients with primary intestinal non-Hodgkin lymphomas based on SEER database [J]. Sci Rep, 2021, 11(1): 23047.https://doi.org/10.1038/s41598-021-02597-1

[25] Lu G, Lin Z, Ruan Y, et al. A Novel Prognostic Model for Patients with Primary Gastric Diffuse Large B-Cell Lymphoma [J]. J Oncol, 2022, 2022: 9636790.https://doi.org/10.1155/2022/9636790

TABLES AND FIGURES WITH LEGENDS

Table 1. Comparison of basic characteristics between the train group and the validation group

Variables	Total (*n* = 4437)	Validation (*n* = 1332)	train (*n* = 3105)	Statistic	*p*
Sex, *n* (%)				χ^2=0.00	0.997
Female	1779 (40.09)	534 (40.09)	1245 (40.10)		
Male	2658 (59.91)	798 (59.91)	1860 (59.90)		
Race, *n* (%)				χ^2=0.90	0.639
Black	289 (6.51)	87 (6.53)	202 (6.51)		
others	584 (13.16)	185 (13.89)	399 (12.85)		
White	3564 (80.32)	1060 (79.58)	2504 (80.64)		
Radiotherapy, *n* (%)				χ^2=0.00	0.968
None/Unknown	3922 (88.39)	1177 (88.36)	2745 (88.41)		
yes	515 (11.61)	155 (11.64)	360 (11.59)		
Chemotherapy, *n* (%)				χ^2=0.84	0.36
No/Unknown	1352 (30.47)	393 (29.50)	959 (30.89)		
Yes	3085 (69.53)	939 (70.50)	2146 (69.11)		
Marital, *n* (%)				χ^2=3.97	0.41
Divorced	310 (6.99)	107 (8.03)	203 (6.54)		
Married	2536 (57.16)	747 (56.08)	1789 (57.62)		
Separated	35 (0.79)	10 (0.75)	25 (0.81)		

Single	762 (17.17)	236 (17.72)	526 (16.94)		
Widowed	794 (17.89)	232 (17.42)	562 (18.10)		
Ann Arbor stage, *n* (%)				χ^2=0.18	0.98
Stage I	1943 (43.79)	577 (43.32)	1366 (43.99)		
Stage II	1188 (26.77)	360 (27.03)	828 (26.67)		
Stage III	315 (7.10)	96 (7.21)	219 (7.05)		
Stage IV	991 (22.33)	299 (22.45)	692 (22.29)		
Age, *n* (%)				χ^2=1.25	0.264
≤60	1248 (28.13)	390 (29.28)	858 (27.63)		
>60	3189 (71.87)	942 (70.72)	2247 (72.37)		
Primary Site, *n* (%)				χ^2=0.35	0.553
stomach	2262 (50.98)	670 (50.30)	1592 (51.27)		
intestine	2175 (49.02)	662 (49.70)	1513 (48.73)		
Surgery, *n* (%)				χ^2=0.10	0.751
No	2807 (63.26)	838 (62.91)	1969 (63.41)		
Yes	1630 (36.74)	494 (37.09)	1136 (36.59)		
LNR, *n* (%)				χ^2=3.50	0.061
No	3647 (82.20)	1073 (80.56)	2574 (82.90)		
Yes	790 (17.80)	259 (19.44)	531 (17.10)		
Income, *n* (%)				χ^2=1.37	0.241

≤$75000	2643 (59.57)	811 (60.89)	1832 (59.00)
>$75000	1794 (40.43)	521 (39.11)	1273 (41.00)

Table 2. Univariate and multivariate Cox regression analysis of prognosis of GI-DLBCL patients in the training cohort

Variables	Univariate analysis		Multivariate analysis	
	HR (95%CI)*	*p*	HR (95%CI)	*p*
Sex				
Female	1.00 (Reference)			
Male	0.98 (0.90 ~ 1.08)	0.733		
Race				
Black	1.00 (Reference)		1.00 (Reference)	
others	0.77 (0.62 ~ 0.96)	0.018	0.65 (0.52 ~ 0.81)	<.001
White	0.93 (0.78 ~ 1.12)	0.457	0.78 (0.65 ~ 0.94)	0.007
Radiotherapy				
None/Unknown	1.00 (Reference)		1.00 (Reference)	
yes	0.75 (0.65 ~ 0.87)	<.001	0.84 (0.72 ~ 0.98)	0.023
Chemotherapy				
No/Unknown	1.00 (Reference)		1.00 (Reference)	
Yes	0.38 (0.35 ~ 0.42)	<.001	0.37 (0.33 ~ 0.41)	<.001
Marital				
Divorced	1.00 (Reference)		1.00 (Reference)	
Married（2536）	0.95 (0.78 ~ 1.14)	0.56	0.99 (0.82 ~ 1.20)	0.923

Separated 35	0.51 (0.26 ~ 1.01)	0.052	1.04 (0.53 ~ 2.05)	0.912
Single 762	0.85 (0.69 ~ 1.06)	0.152	1.22 (0.98 ~ 1.51)	0.077
Widowed	1.83 (1.49 ~ 2.24)	<.001	1.40 (1.14 ~ 1.72)	0.001
Ann Arbor stage				
Stage I	1.00 (Reference)		1.00 (Reference)	
Stage II	0.94 (0.84 ~ 1.05)	0.267	1.22 (1.08 ~ 1.37)	<.001
Stage III	1.01 (0.84 ~ 1.22)	0.905	1.31 (1.08 ~ 1.58)	0.005
Stage IV	1.52 (1.36 ~ 1.70)	<.001	1.85 (1.65 ~ 2.08)	<.001
Age				
≤60	1.00 (Reference)		1.00 (Reference)	
>60	3.13 (2.75 ~ 3.55)	<.001	2.85 (2.49 ~ 3.26)	<.001
Primary Site				
stomach	1.00 (Reference)		1.00 (Reference)	
intestine	0.85 (0.78 ~ 0.93)	<.001	0.89 (0.79 ~ 0.99)	0.037
Surgery				
No	1.00 (Reference)		1.00 (Reference)	
Yes	0.89 (0.81 ~ 0.98)	0.015	0.86 (0.75 ~ 0.99)	0.032
LNR				
No	1.00 (Reference)		1.00 (Reference)	
Yes	0.84 (0.74 ~ 0.95)	0.004	0.79 (0.68 ~ 0.92)	0.002
Income				
≤75000	1.00 (Reference)			
>75000	0.91 (0.83 ~ 1.00)	0.056		

*HR (95%CI): Hazard Ratio (HR), 95% (Confidence Interval, CI), HR is within the CI range with 95% probability.

Note: Hazard Ratio (HR), HR>1 is a risk factor (such as age>60 years, Stage IV, Widowed), HR<1 is a protective factor (such as chemotherapy, primary site at intestine), HR=1 is no effect (such as black ethnicity in multiple factors). 95% CI represent HR value is within the CI range with 95% probability, which excluding 1 indicate results are statistically significant, and including 1 indicate results are not statistically significant. The HR=1 (reference) are defined as reference values for comparison with independent risk factors. Probability value (P) <0.05 suggest statistically significant, P ≥ 0.05 suggest not statistically significant.

Table 3. C-index of the nomogram model and Ann Arbor staging system

Classification	Training group		Validation group	
	C-index (95%CI)	*p* value	C-index (95%CI)	*p* value
Nomogram	0.71 (0.70-0.73)	1	0.71 (0.69-0.73)	1
Ann Arbor stage	0.56 (0.54-0.57)	<0.001	0.56 (0.54-0.57)	<0.001
ΔC-index (Nomogram - Ann Arbor)	0.15	/	0.15	/

Table 4. AUC of the nomogram model and Ann Arbor staging system

Classification	Training group AUC (95%CI)	Validation group AUC (95%CI)
Nomogram (3 years)	0.749 (0.71-0.81)	0.753 (0.72-0.82)
Nomogram (5 years)	0.753 (0.72-0.84)	0.744 (0.68-0.80)
Nomogram (10 years)	0.755 (0.73-0.88)	0.737 (0.65-0.79)
Ann Arbor stage (3 years)	0.565 (0.52-0.63)	0.610 (0.58-0.66)
Ann Arbor stage (5 years)	0.561 (0.50-0.62)	0.592 (0.54-0.65)
Ann Arbor stage (10 years)	0.561 (0.49-0.62)	0.567 (0.52-0.64)

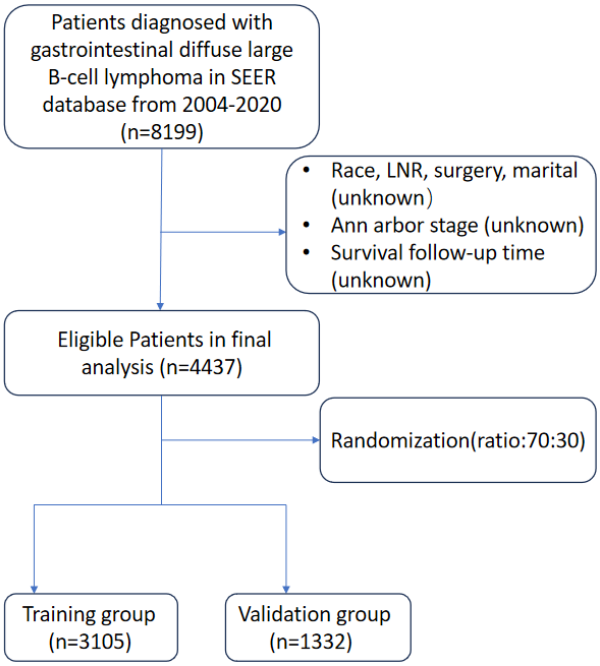

Figure 1. Flow chart of patient inclusion

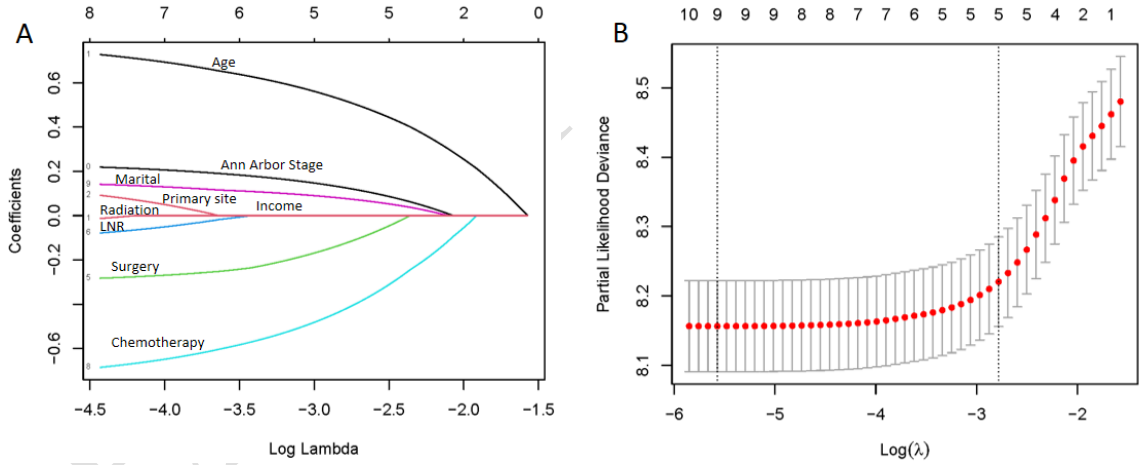

Figure 2. LASSO regression analysis of different variables. (A)The coefficient path plot of variable coefficients as a function of Log λ (where λ is the penalty intensity), reflecting the contraction process of variable coefficients β (vertical axis) under different penalty intensities. Positiveβ indicate a positive correlation between variables and risk factors; (B) Partial likelihood deviance under cross validation, which is used to select the optimal value of λ.

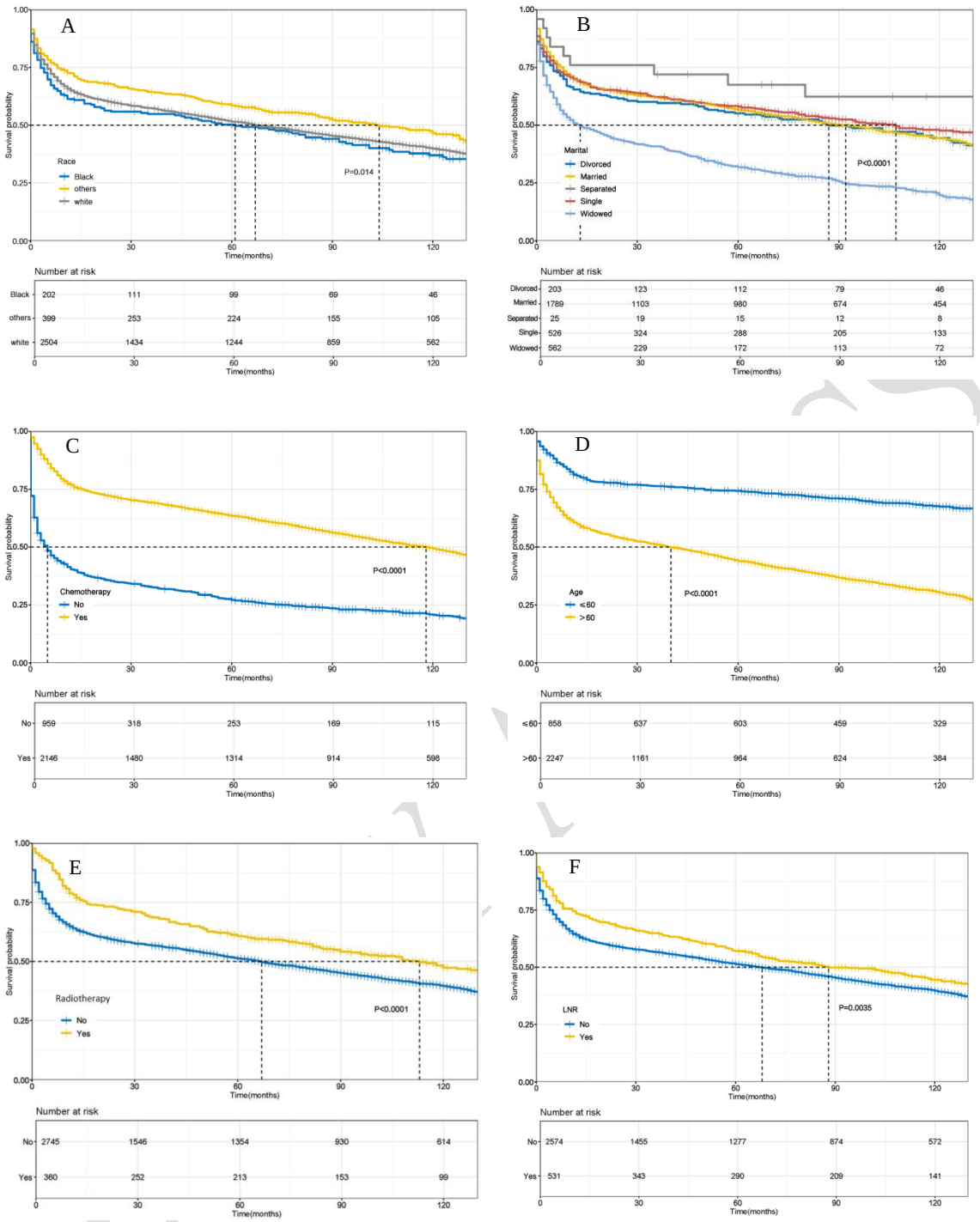

Figure 3. Kaplan–Meier survival curves for six independent prognostic factors in patients with GI-DLBCL.Vertical dashed lines represent median follow-up time. **(A)**: Race; **(B)**: Marital status; **(C)**: Chemotherapy; **(D)**: Age; **(E)**: Radiotherapy; **(F)**: Lymph node resection (LNR).

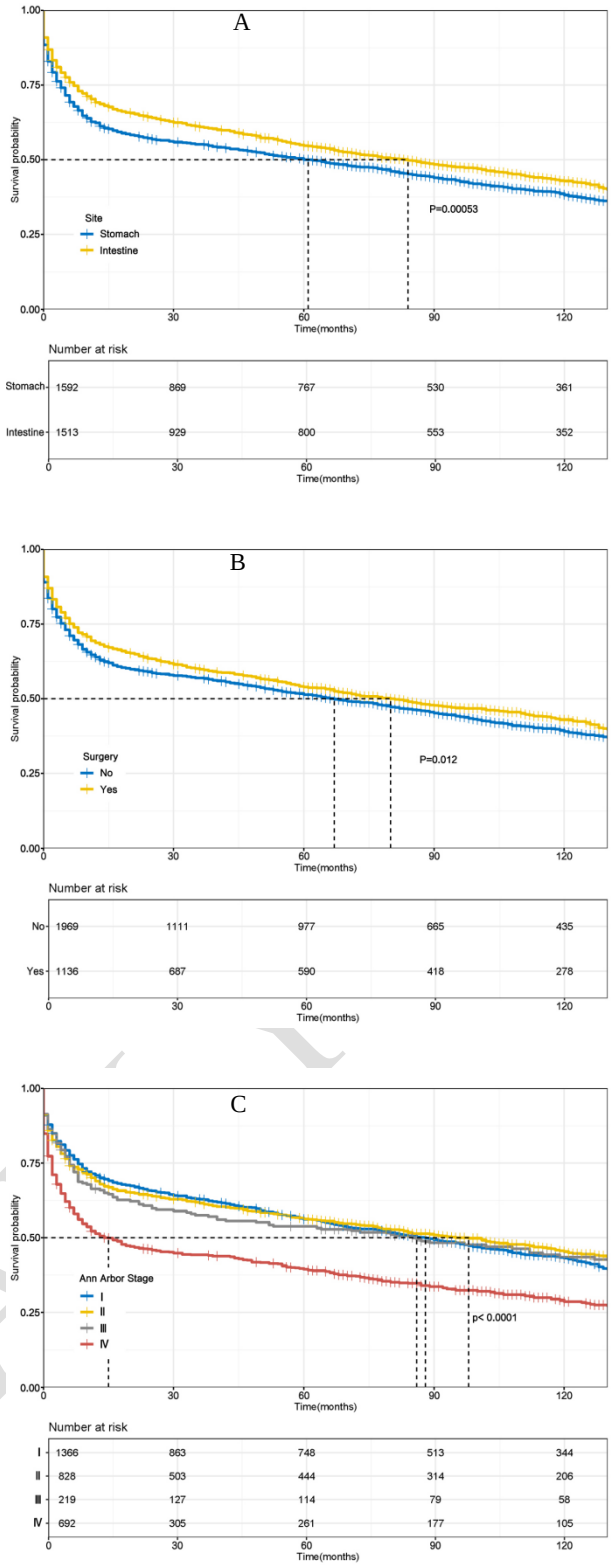

Figure 4. Kaplan–Meier survival curves for three independent prognostic factors in patients with GI-DLBCL. Vertical dashed lines represent median follow-up time. **(A)**: Primary site; **(B)**: Surgery; **(C)**: Ann Arbor stage.

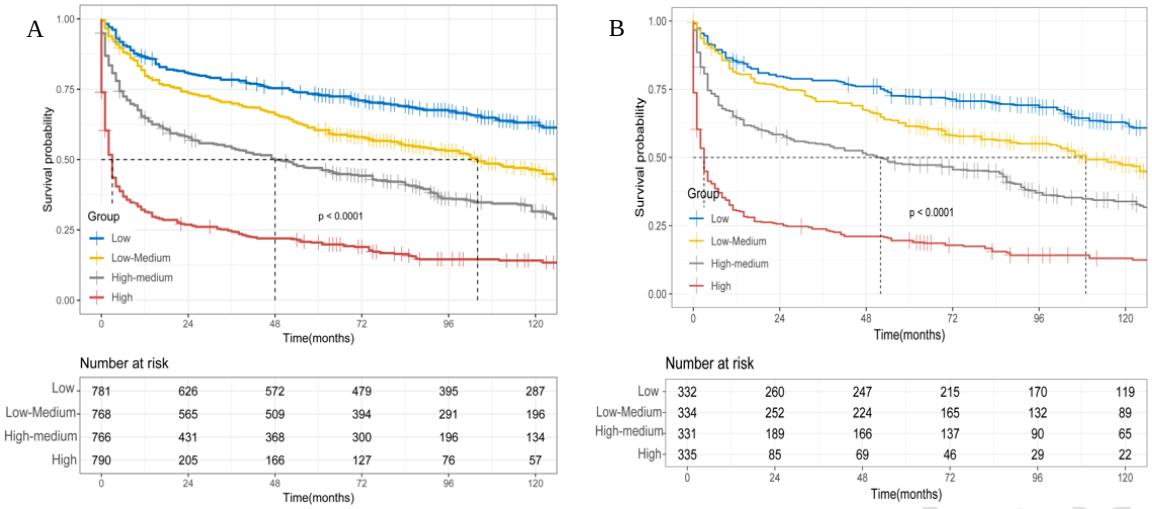

Figure 5. All the multivariate factors related Kaplan-Meier hierarchical survival curves. **(A)**Training group;**(B)** Validation group.

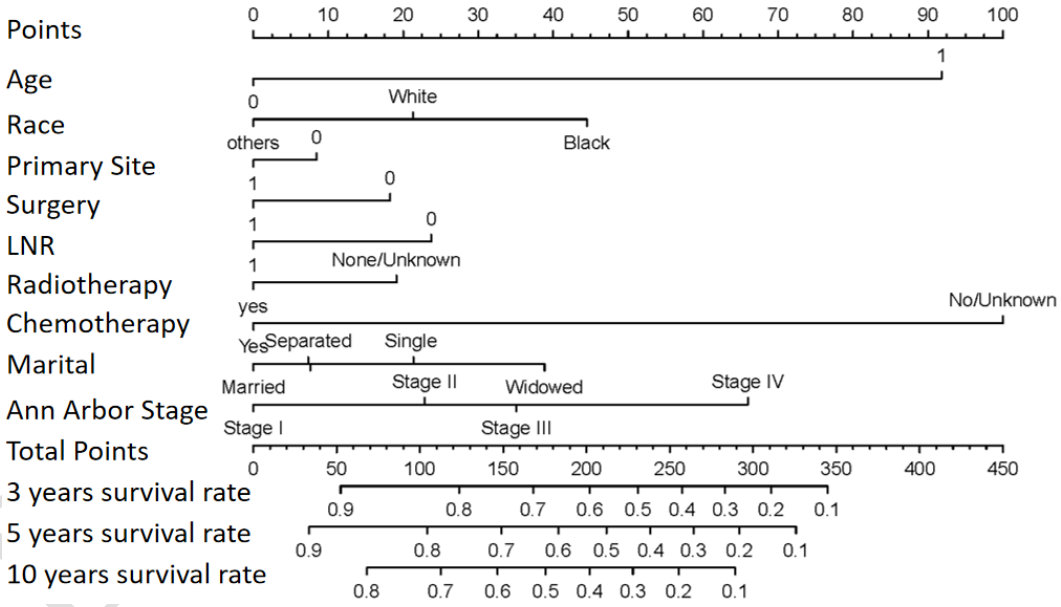

Figure 6. Nomogram for predicting 3-, 5-, and 10-year overall survival in patients with GI-DLBCL

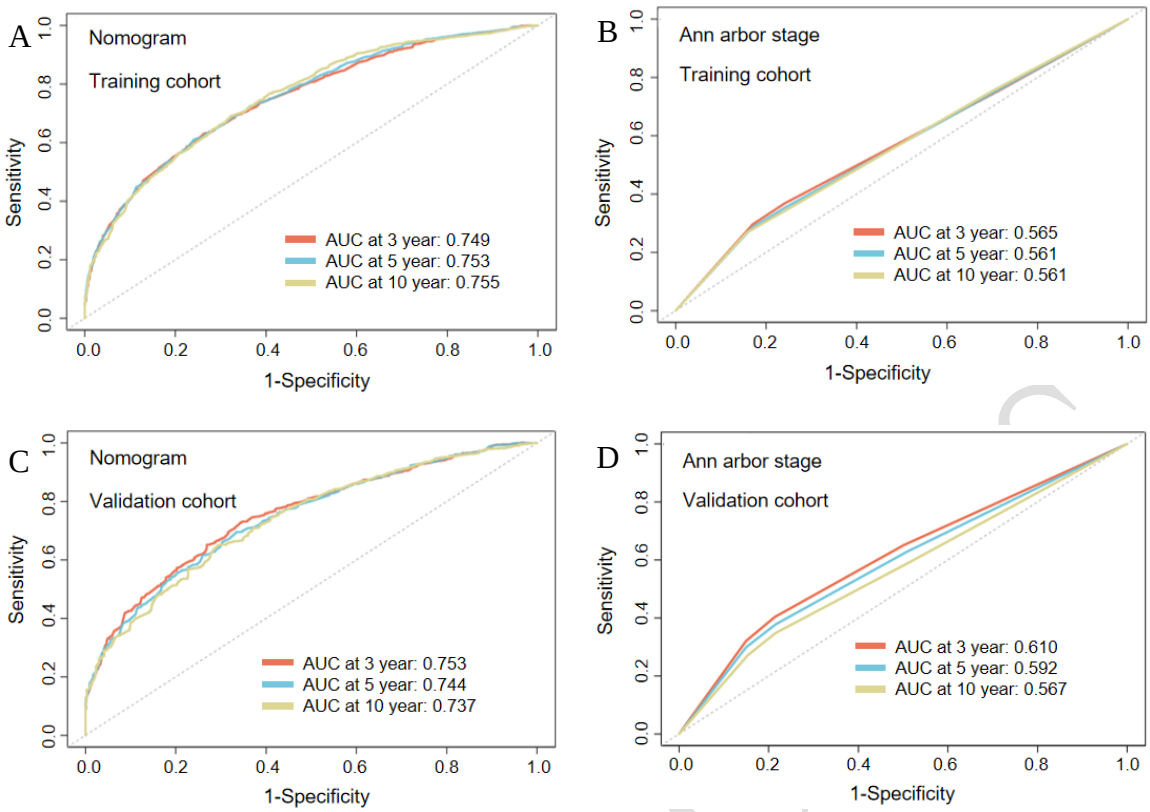

Figure 7. ROC curve analysis of the prognostic predictive performance of the nomogram model versus the Ann Arbor staging system in patients with GI-DLBCL. **(A–B)**: ROC curves and AUC values for predicting 3-, 5-, and 10-year overall survival in GI-DLBCL patients from the training cohort, where **(A)** represents the nomogram model and **(B)** represents the Ann Arbor staging System;**(C–D)**: ROC curves and AUC values for predicting 3-, 5-, and 10-year overall survival in GI-DLBCL patients from the internal validation cohort, where **(C)** represents the nomogram model and **(D)** represents the Ann Arbor staging System.

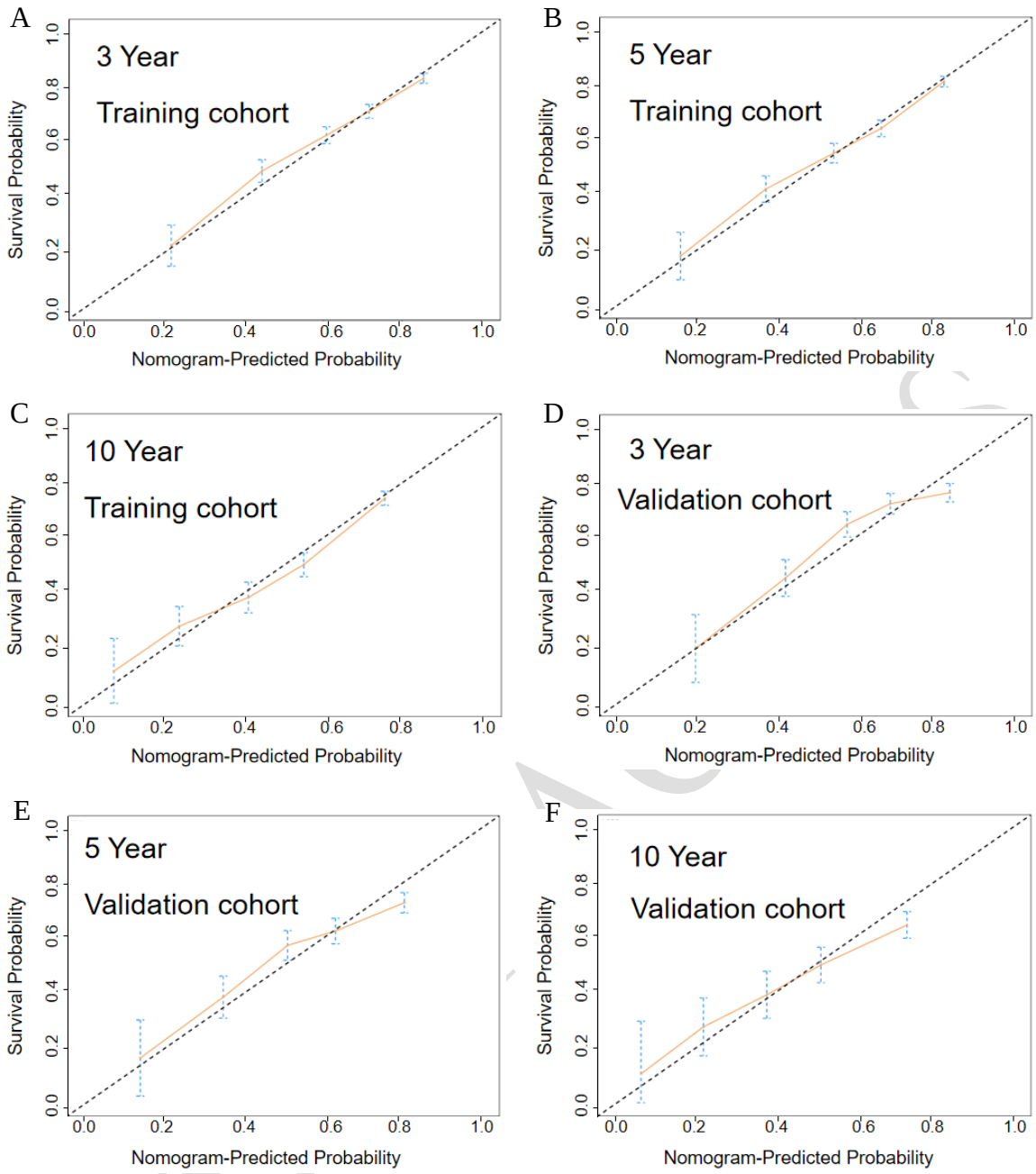

Figure 8. Calibration curve analysis of the nomogram model for predicting overall survival in patients with GI-DLBCL. **(A–C)**: Calibration curves for predicting 3-, 5-, and 10-year overall survival in GI-DLBCL patients from the training cohort; **(D–F)**: Calibration curves for predicting 3-, 5-, and 10-year overall survival in GI-DLBCL patients from the internal validation cohort.

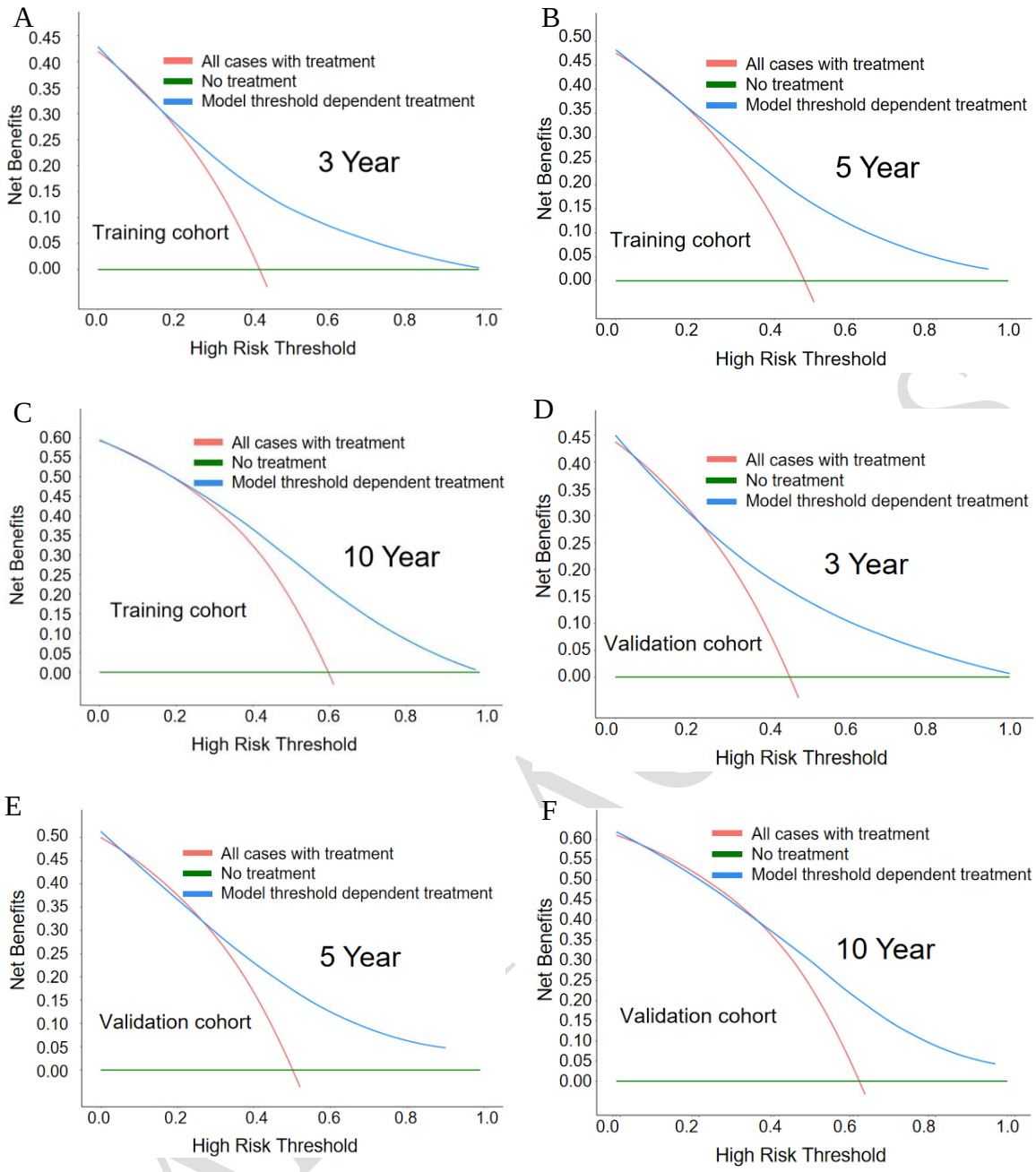

Figure 9. Decision curve analysis (DCA) of the Nomogram model for clinical utility in predicting survival of patients with GI-DLBCL. **(A–C)**: DCA for the 3-, 5-, and 10-year survival prediction models in GI-DLBCL patients from the training cohort;**(D–F)**: DCA for the 3-, 5-, and 10-year survival prediction models in GI-DLBCL patients from the internal validation cohort.

SUPPLEMENTAL DATA

Supplemental data are available at the following link:

https://www.bjbms.org/ojs/index.php/bjbms/article/view/12697/3956