

Bladder cancer survival nomogram

Development and validation of a prediction tool, using the SEER and TCGA databases

Ye Zhang, MS, Ying-kai Hong, MS, Dong-wu Zhuang, MS, Xue-jun He, MS, Ming-en Lin, MD, PhD*

Abstract

Bladder cancer (BC) is a common malignancy associated with high morbidity and mortality, however, accurate and convenient risk assessment tools applicable to BC patients are currently lacking. Previous studies using nomograms to evaluate bladder cancer (BC) survival have been based on small samples. Using a large dataset, this study aimed to construct more precise clinical nomograms to effectively predict bladder cancer survival.

Data on patients with pathologically-confirmed bladder cancer were extracted from the Surveillance, Epidemiology, and End Results (SEER) database. Additional BC patient data for an external validation cohort were extracted from the Cancer Genome Atlas (TCGA) database. Clinical parameters that constituted potential risk factors were reviewed and analyzed using univariate and multivariate Cox proportional hazards regression. A nomogram was constructed with parameters that significantly correlated with the overall survival (OS). Prognostic performance of a nomogram was assessed using the concordance index (c-index), area under the receiver operating characteristic curve (AUC), and a calibration curve. The model was then tested with data from an internal and external validation cohort. Patients' survival was analyzed and compared with the Kaplan-Meier (KM) method.

Multivariate Cox regression showed that age, sex, race, stage_T1, stage_T2a, stage_T2b, stage_T3a, stage-Ta, stage_Tis, stage_N, stage_M were independent predictors of BC survival. A nomogram was constructed based on these factors. The c-index of the nomogram was 0.7916 (95% confidence interval CI, 0.79–0.80). The calibration curve showed excellent agreement between the predicted and observed values. The c-index for the internal validation cohort was 0.7917 (95% CI 0.79–0.80), which was higher than for the training cohort, suggesting robustness of the model. For the training cohort, the AUC for the 3- and the 5-year survival was 0.82 and 0.813, respectively. The c-index for the TNM-based model was superior to that for the AJCC-TNM classification.

The models presented in this study might be suitable for clinical use, supporting clinicians in their individualized assessment of expected survival in BC patients. They might also be used as a layered tool for clinical research.

Abbreviations: AJCC = American Joint Committee on Cancer, AUC = the area under the time-dependent receiver operating characteristic curve, BC = bladder cancer, C-index = concordance index, HR = hazard ratio, NMIBC = non-muscle invasive bladder cancer, OS = overall survival, ROC = receiver operating characteristic curve, SEER = surveillance, epidemiology, and end result, TCGA = the cancer genome atlas, TNM = tumor node metastasis, US = United States.

Keywords: bladder cancer, data-driven model, nomogram, overall survival, TNM, validation

Editor: Giuseppe Lucarelli.

This work was supported by Natural Science Foundation of Guangdong province (2015A030310078), Guangdong Medical Research Foundation (A2018103) and Shantou Science and Technology Fund (2017026).

The authors report no conflicts of interest

Supplemental Digital Content is available for this article.

Department of Urology, the First affiliated hospital of Medical College Shantou University, 57 Changping Road, Shantou, Guangdong, China.

* Correspondence: Ming-en Lin, Department of Urology, the First affiliated hospital of Medical College Shantou University, 57 Changping Road, Shantou 515000, Guangdong, China (e-mail: me_lin20@163.com).

Copyright © 2019 the Author(s). Published by Wolters Kluwer Health, Inc. This is an open access article distributed under the terms of the Creative Commons Attribution-Non Commercial License 4.0 (CCBY-NC), where it is permissible to download, share, remix, transform, and buildup the work provided it is properly cited. The work cannot be used commercially without permission from the journal.

How to cite this article: Zhang Y, Hong Yk, Zhuang Dw, He Xj, Lin Me. Bladder cancer survival nomogram: Development and validation of a prediction tool, using the SEER and TCGA databases. *Medicine* 2019;98:44(e17725).

Received: 5 July 2019 / Received in final form: 25 September 2019 / Accepted: 25 September 2019

<http://dx.doi.org/10.1097/MD.0000000000017725>

1. Introduction

Bladder cancer (BC) is a common urinary malignant tumor,^[1] characterized by high morbidity and mortality. In 2018 in the United States (US) alone, there were 81,190 newly-diagnosed BC cases, and 17,240 BC-attributable deaths.^[2] Approximately 25% of BC patients present with muscle-invasive BC or metastatic disease, while 75% present with non-muscle invasive BC (NMIBC).^[3] The proportion of patients with NMIBC is relatively high; however, the high rate of recurrence (70%) in low- and intermediate-risk disease, and the fairly high rate of progression to muscle-invasive disease (30%) in high-risk NMIBC are cause for concern.^[4–6] The majority of BC cases occur in people aged over 60. The main risk factor for BC is increasing age, but smoking and exposure to some industrial chemicals have also been reported as risk factors.^[7]

Numerous staging systems have been proposed for urinary bladder carcinoma and the most commonly used is the American Joint Committee on Cancer (AJCC) staging system.^[8] The eighth edition of the US Joint Commission on Cancer (AJCC) staging manual, has established a tumor node metastasis (TNM) classification system, which indicates increasing understanding

of the pathophysiology of BC and applicable treatments. This staging system is currently widely used in making prognostic estimates and treatment decisions in clinical practice.^[9] However, there is evidence to suggest that patients with the same pathological grade or clinical staging of BC might still have different prognosis and ultimate survival. This further suggests existence of factors not directly related to the characteristics of BC that might affect patients' prognosis. To predict survival accurately and reliably, the use of nomograms has been proposed. A method based on more refined TNM staging for predicting individualized survival of bladder cancer patients is required, and a nomogram is a good method for this purpose. Nomograms are based on the TNM staging system, and other key prognostic factors associated with patient survival, and have been applied as a layered tool in clinical research on prostate, breast, gastric, and colorectal cancer.^[10–13] Analysis of these nomograms has led to identifying potential survival prognostic factors. Overall, using the c-index, these models are considered to surpass clinical judgment when predicting patient survival.^[10,14] Within BC research, previous studies have found potential prognostic factors through the construction of novel nomograms; for example, age, stage, grade,^[15] tumor size, lymphovascular invasion, variant histology,^[16] genetic variants (smad6, fn1, galectin-9, p53, pRB, p21, p27, and cyclin E1),^[17–19] and hemocyte type (hemoglobin, albumin, lymphocyte, and platelet type) have been identified using this approach.^[20] However, to assess accurately the prognosis of BC patients, studies involving a large number of patients, and internal and external validation datasets, are required. To our knowledge, this study is the first attempt to construct a clinical nomogram for bladder cancer survival, using a large database. We further construct a population-based survival-predicting model with internal validations. In addition, we used a TCGA cohort to externally validate the nomogram.

2. Materials and methods

2.1. Patients and study design

We identified BC cases from the Surveillance, Epidemiology, and End Results (SEER) database of the National Cancer Institute (<http://seer.cancer.gov/>) before we signed Research Data Agreement is on file at SEER. We are allowed to utilize the SEER*Stat client-server system and/or download the files which make up the SEER Research Data. To be included in the study, patients had to have had pathologically-confirmed BC recorded in one of the 18 SEER-covered registries at any point during the database's coverage (1973–2015). We extracted data on the patients' age, sex, race, stage (T/N/M), survival time, and mortality. Patients were excluded if any of the data was missing or incomplete. We used an internal verification method to randomly divide our dataset into 2 cohorts (training and a validation cohort at ratio 7:3). In addition, we also extracted data on 130 patients from the Cancer Genome Atlas (TCGA) (<https://portal.gdc.cancer.gov/>), which is a publicly available database, using as an external validation cohort. The clinical information of BC is publicly available in the SEER and TCGA program, so the approval of local ethics committee was not needed.

2.2. Statistical analysis

First, univariate Cox regression analysis including selected parameters was carried out. Second, independent prognostic

factors were identified in multivariate Cox proportional hazards regression analysis. Third, a nomogram based on these prognostic factors was constructed base on the training cohort data.

The Cox proportional hazard regression model was used to estimate the hazard ratio (HR), and the corresponding 95% confidence interval (CI), for each of the potential risk factors. The multivariate Cox regression model was constructed with a nomogram predicting 3- and 5-year BC survival. Validation of the nomogram was performed using the c-index, the area under the receiver operating characteristic curve (AUC), and the calibration curve. The nomogram-predicted survival probability was compared with observed survival probability, calculated with the Kaplan–Meier (KM) method. The c-index was used to estimate the predictive accuracy and discrimination ability of each factor as well as of the overall nomogram: the higher the c-index, the better its prognostic accuracy. The receiver operating characteristics (ROC) curves are similar to the c-index, but are considered less suitable for use with censored data. The calibration curves were used to assess the nomogram-predicted 3- and 5-year survival with the observed 3- and 5-year survival. Patients were divided into high-risk and low-risk with the cut-off value set at the median risk, estimated in the KM analysis. To achieve this, we estimated an optimism-corrected calibration curve with a bootstrapped sample of 1000. Data extraction was performed using the SEER*Stat software version 8.3.5. Statistical analyses were performed using R software version 3.5.3 (<http://www.r-project.org>) with the RMS, survival, and foreign statistical packages. For all of the analyses, a two-tailed *P* value < .05 was considered statistically significant.

3. Results

3.1. Patient characteristics

Data on a total of 398,173 patients were extracted from the SEER database, according to the screening criteria. Subsequently, data from 332,202 patients were excluded, as they had not been assigned accurate parameters. The final sample included 65,971 patients in the entire cohort (Table 1). Subsequently, the cohort was divided and a total of 46,179 (70%) patients were included in the training cohort, while 19,792 (30%) patients were included in the internal validation cohort. The external validation cohort included 130 patients from the TCGA (Table 2). The clinicopathological characteristics of the training and validation cohorts are shown in Table 1. The median survival time for BC patients in this sample was 38 months, with the specific 3- and 5-year survival rates at 52% and 13%, respectively (Fig. 1) in the SEER cohort.

3.2. Independent bladder cancer survival prognostic factors

Univariate Cox regression analysis of the training cohort revealed a role of the following parameters in predicting patient survival. Factors such as age, sex, race, stage_T1, stage_T2a, stage_T2b, stage_T2NOS, stage_T3a, stage_T3b, stage_T3NOS, stage_T4a, stage_Ta, stage_Tis, stage_TX, stage_N, stage_M were associated with patients' prognosis. Among these factors, stage_T (c-index=0.729) and age (c-index=0.645) each had superior discrimination power in predicting BC survival compared with other factors. The results of multivariate analyses,

Table 1
Characteristics of 65,971 patients with Bladder Cancer in SEER.

Factors	All cohort	%	Traincohort	%	Validationcohort	%
Age						
<50	2362	3.5	1671	3.6	691	3.4
50–59	7302	11	5128	11.1	2174	10.9
60–69	16,478	24.9	11,482	24.8	4996	25.2
70–79	19,724	29.8	13,881	30	5843	29.5
>80	20,105	30.4	14,017	30.3	6088	30.7
Race						
Black	3431	5.2	2378	5.1	1053	5.3
White	58,476	88.6	40,954	88.6	17,522	88.5
Others	4064	6.1	2847	6.1	1217	6.1
Sex						
Female	15,469	23.4	10,806	23.4	4663	23.5
Male	50,502	76.5	35,373	76.5	15,129	76.4
Stage_T						
stage_T0	31	0.05	20	0.04	11	0.05
stage_T1	16,068	24.3	11,190	24.2	4878	24.6
stage_T2a	2981	4.5	2131	4.6	850	4.2
stage_T2b	1727	2.6	1174	2.5	553	2.7
stage_T2NOS	4850	7.3	3405	7.3	1445	7.3
stage_T3a	1447	2.1	1015	2.1	432	2.1
stage_T3b	819	1.2	570	1.2	249	1.2
stage_T3NOS	454	0.6	337	0.7	117	0.5
stage_T4a	1844	2.7	1290	2.7	554	2.7
stage_T4b	441	0.6	309	0.6	132	0.6
stage_T4NOS	33	0.05	26	0.05	7	0.04
stage_Ta	29,976	45.4	20,984	45.4	8992	45.4
stage_Tis	3059	4.6	2155	4.6	904	4.5
stage_TX	2241	3.3	1573	3.4	668	3.3
Stage_N						
stage_N0	60,806	92.1	42,540	92.1	18,266	92.2
stage_N1	1116	1.6	815	1.7	301	1.5
stage_N2	1468	2.2	1019	2.2	449	2.2
stage_N3	365	0.5	249	0.5	116	0.5
stage_NX	2216	3.3	1556	3.3	660	3.3
Stage_M						
stage_M0	63,373	96	44,402	96.1	18,971	95.8
stage_M1	2598	3.9	1777	3.8	821	4.1

Table 2
Characteristics of 130 patients with Bladder Cancer in TCGA.

Factors	Cohort	%
age		
<50	3	2.30%
50–59	20	15.30%
60–69	42	32.30%
70–79	50	38.40%
>>80	15	11.50%
race		
Black	12	9.20%
White	106	81.50%
Others	12	9.20%
sex		
Female	27	20.70%
Male	103	79.20%
stage_T		
stage_T1	3	2.30%
stage_T2a	17	13%
stage_T2b	23	17.60%
stage_T3a	26	20%
stage_T3b	38	29.20%
stage_T4a	17	13%
stage_T4b	3	2.30%
stage_TX	3	2.30%
stage_N		
stage_N0	83	63.80%
stage_N1	14	10.70%
stage_N2	18	13.80%
stage_N3	2	1.50%
stage_NX	13	10%
stage_M		
stage_M0	120	92.30%
stage_M1	10	7.60%

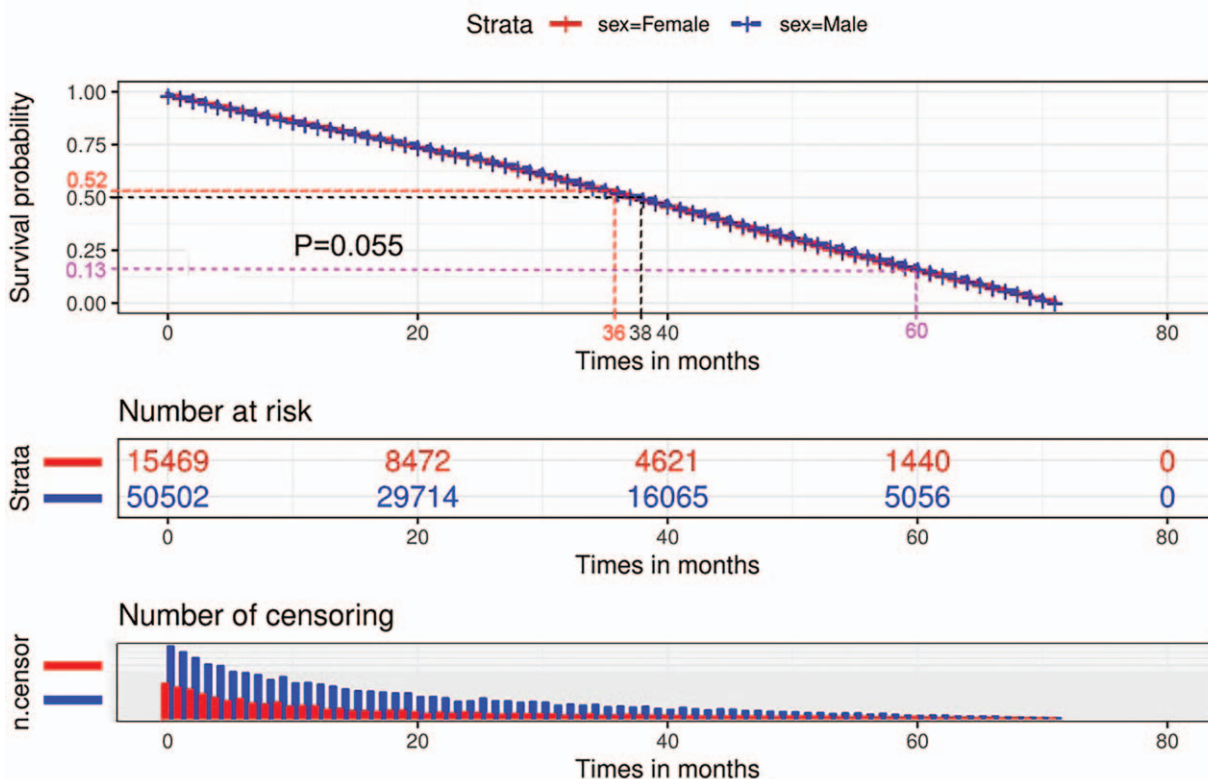


Figure 1. Kaplan–Meier survival plot for gender.

Table 3
Overall survival stratified by clinical characteristics and univariate Cox proportional hazard analyses for patients with Bladder Cancer.

	exp (coef)	lower.95	upper.95	P
age < 50 reference				
age50–59	1.320	1.134	1.538	.000358***
age60–69	1.603	1.390	1.850	9.2e–11***
age70–79	2.476	2.153	2.849	< 2e–16***
age ≥ 80	5.412	4.712	6.217	< 2e–16***
raceBlack reference				
raceWhite	0.7289	0.6807	0.7807	< 2e–16***
raceOther	0.6955	0.6314	0.7660	1.76e–13***
sexFemale reference				
sexMale	0.8459	0.8144	0.8787	< 2e–16***
stage_T0 reference				
stage_TT1	0.11436	0.06997	0.18692	< 2e–16***
stage_TT2a	0.25898	0.15810	0.42424	8.09e–08***
stage_TT2b	0.26935	0.16399	0.44240	2.20e–07***
stage_TT2NOS	0.32532	0.19888	0.53212	7.72e–06***
stage_TT3a	0.25153	0.15295	0.41363	5.38e–08***
stage_TT3b	0.33250	0.20142	0.54891	1.67e–05***
stage_TT3NOS	0.34823	0.20939	0.57912	4.81e–05***
stage_TT4a	0.45821	0.27947	0.75129	.00198**
stage_TT4b	0.89137	0.53805	1.47671	.65525
stage_TT4NOS	0.69021	0.36018	1.32267	.26389
stage_TTa	0.05343	0.03269	0.08734	< 2e–16***
stage_TTis	0.07932	0.04820	0.13053	< 2e–16***
stage_TTX	0.48841	0.29809	0.80025	.00445**
stage_N0 reference				
stage_NN1	3.541	3.244	3.865	< 2e–16 ***
stage_NN2	4.990	4.633	5.374	< 2e–16 ***
stage_NN3	5.263	4.553	6.083	< 2e–16 ***
stage_NNX	4.151	3.892	4.427	< 2e–16 ***
stage_M0 reference				
stage_MM1	9.153	8.667	9.666	< 2e–16 ***

Signif.codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1.

with stepwise models including risk factors identified as significant in univariate analysis, showed that age, sex, race, stage_T1, stage_T2a, stage_T2b, stage_T3a, stage_Ta, stage_Tis, stage_N, and stage_M were independent predictors of BC survival (Tables 3 and 4). These factors were subsequently included in the predictive model. In the analysis of the TCGA data used for external validation, only lymph node metastasis was associated with survival (Table 5).

3.3. Prognostic nomogram for OS

Predictive models with nomograms integrating all factors affecting survival in the training cohort are shown in Figure 2. Each prognostic parameter was assigned a score according to its prognostic value; the sum total of the scores was used to predict 3- and 5-year survival. The total score for all the variables was converted into an estimate of the probability of death. The c-index of the prognostic nomogram for overall survival prediction was 0.7916 (95% CI, 0.79–0.80) in the training cohort, and 0.7917 (95% CI, 0.79–0.80) in the internal validation cohort. The validation set was superior to the training set, indicating the robustness of the model. The AUCs for the 3- and 5-year survival were 0.82 and 0.813, respectively, in training cohort (Fig. 3A, B). The AUC combined with the c-index reflected good discrimination ability of the model. The calibration curves estimating

Table 4
Multivariate Cox proportional hazard analyses of clinical characteristics for overall survival rates in patients with Bladder Cancer.

	exp (coef)	lower.95	upper.95	P
age < 50 reference				
age50–59	1.3263	1.13838	1.5451	.000291***
age60–69	1.6607	1.43935	1.9160	3.64e–12***
age70–79	2.6101	2.26814	3.0035	< 2e–16***
age ≥ 80	5.7057	4.96495	6.5570	< 2e–16***
sexFemale reference				
sexMale	0.9279	0.89299	0.9641	.000128***
raceBlack reference				
raceWhite	0.7532	0.70298	0.8071	8.86e–16***
raceOther	0.6590	0.59808	0.7261	< 2e–16***
stage_T0 reference				
stage_TT1	0.2496	0.15222	0.4094	3.82e–08***
stage_TT2a	0.5236	0.31868	0.8602	.010639*
stage_TT2b	0.5422	0.32922	0.8931	.016204*
stage_TT2NOS	0.6197	0.37786	1.0164	.058015
stage_TT3a	0.5439	0.32940	0.8980	.017286*
stage_TT3b	0.6899	0.41633	1.1432	.149673
stage_TT3NOS	0.6602	0.39591	1.1009	.111481
stage_TT4a	0.7820	0.47557	1.2858	.332458
stage_TT4b	1.3739	0.82758	2.2808	.219375
stage_TT4NOS	0.9451	0.49275	1.8128	.865132
stage_TTa	0.1295	0.07892	0.2124	5.71e–16***
stage_TTis	0.1823	0.11039	0.3010	2.91e–11***
stage_TTX	0.6820	0.41559	1.1192	.129943
stage_NN0 reference				
stage_NN1	1.3689	1.24801	1.5014	2.77e–11***
stage_NN2	1.5867	1.46115	1.7230	< 2e–16***
stage_NN3	1.3899	1.19475	1.6168	1.99e–05***
stage_NNX	1.3803	1.27373	1.4957	3.75e–15***
stage_MM0 reference				
stage_MM1	3.6626	3.44415	3.8950	< 2e–16***

Signif.codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1.

Table 5
Multivariate Cox proportional hazard analyses of clinical characteristics for overall survival rates in TCGA cohort.

	exp (coef)	lower.95	upper.95	P
age < 50 reference				
50–59	1.274e+00	0.13732	11.821	.83123
60–69	1.311e+00	0.15809	10.878	.80171
70–79	3.057e+00	0.36801	25.393	.30091
> = 80	1.831e+00	0.19702	17.020	.59484
sexFemale reference				
Male	6.083e–01	0.31598	1.171	.13691
raceBlack reference				
White	8.342e–01	0.31649	2.199	.71387
Other	8.142e–01	0.18892	3.509	.78274
stage_T reference				
stage_T2a	3.289e+07	0.00000	Inf	.99669
stage_T2b	6.417e+07	0.00000	Inf	.99656
stage_T3a	1.103e+08	0.00000	Inf	.99645
stage_T3b	7.593e+07	0.00000	Inf	.99653
stage_T4a	9.053e+07	0.00000	Inf	.99649
stage_T4b	2.878e+08	0.00000	Inf	.99627
stage_TX	3.730e+08	0.00000	Inf	.99622
stage_N reference				
stage_N1	3.470e+00	1.53213	7.861	.00286**
stage_N2	2.580e+00	1.07648	6.184	.03357*
stage_N3	6.527e–01	0.06174	6.901	.72294
stage_NX	2.843e+00	1.04050	7.766	.04161*
stage_M reference				
stage_M1	1.894e+00	0.64815	5.534	.24308

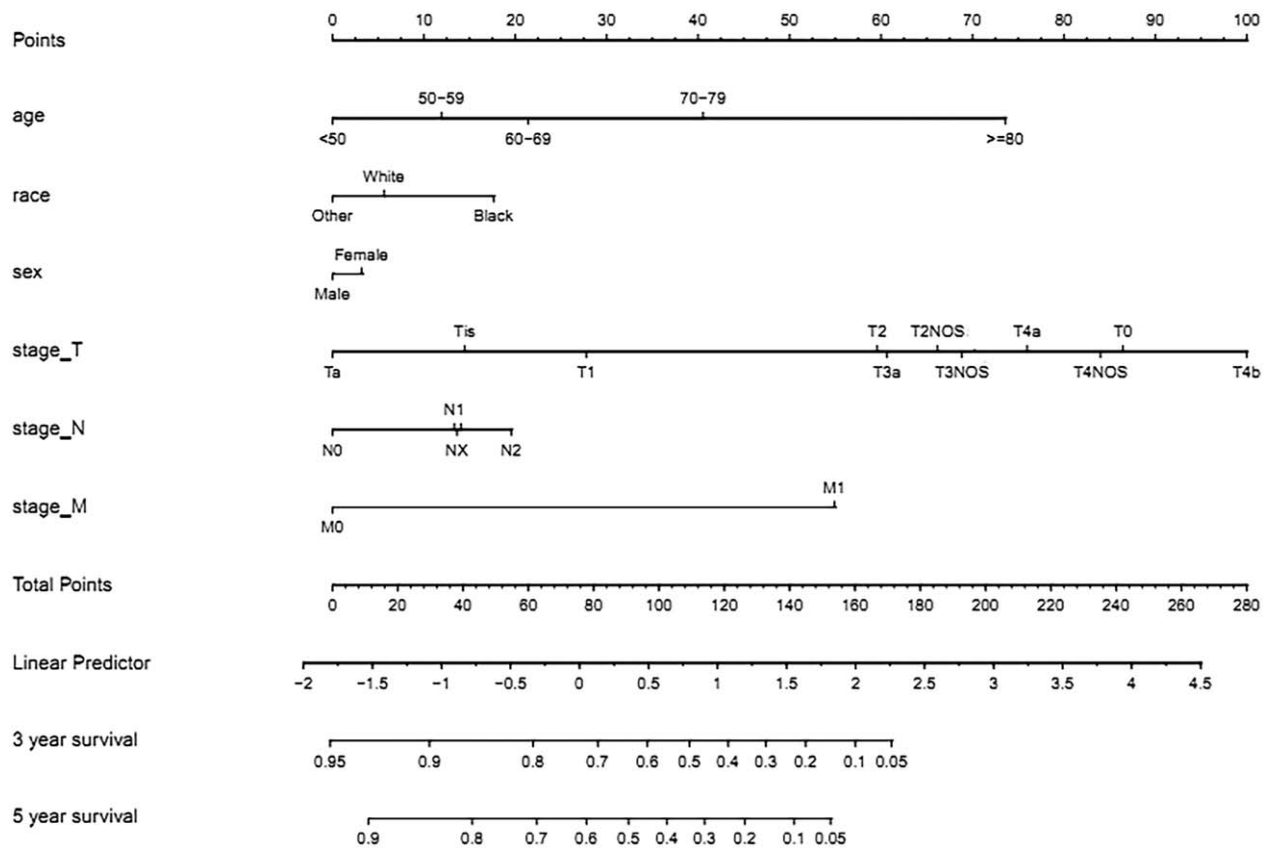


Figure 2. Nomogram predicting the 3-year and 5-year OS for bladder cancer patients. The nomogram summed the points identified on the scale for each variable. The total points projected on the bottom scales indicate the probabilities of 3- and 5-year overall survival.

survival probability at 3 and 5 years showed excellent agreement between the nomogram-predicted and observed values (Fig. 4A, B). When patients were divided into high-risk and low-risk groups, with median risk used as the cutoff point, the survival

curves showed significant differences in prognosis. The 3- and 5-year survival rates for the low-risk group were 85% and 80%, respectively; the same outcomes for the high-risk group were 45% and 35%, respectively (Fig. 5A). Age was strongly

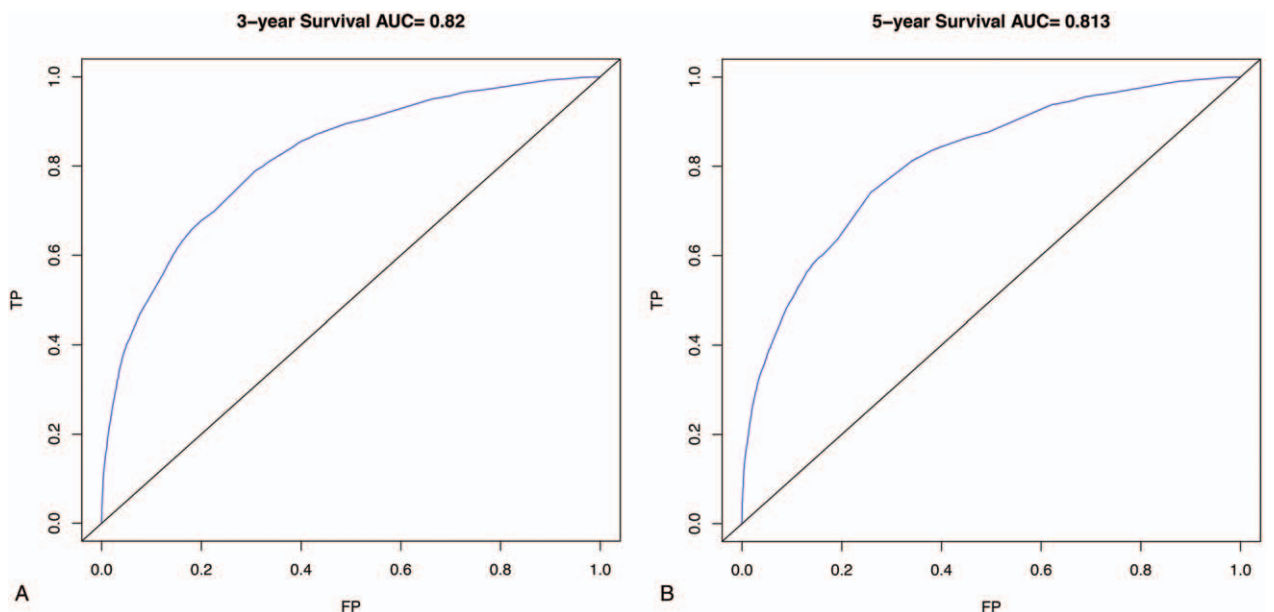


Figure 3. Area under the curves to predict overall survival at 3 years (A) and 5 years (B) using train cohort.

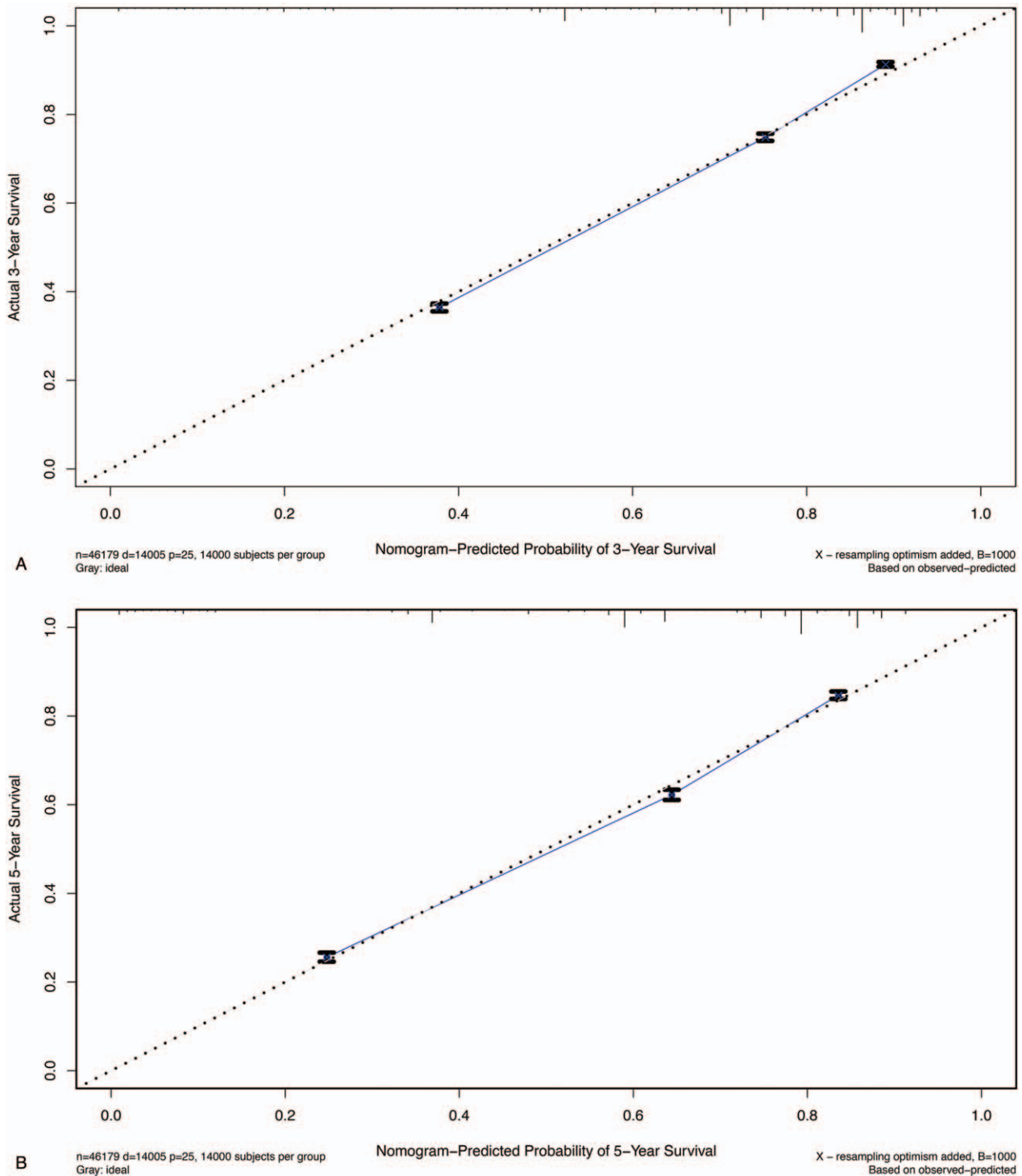


Figure 4. The calibration curves for predicting patient survival at (A) 3-year and (B) 5-year in the train cohort. Nomogram-predicted survival is plotted on the x-axis, and the actual survival is plotted on the y-axis.

predictive of BC survival (Fig. 5B), as was race (Fig. 5C). The 3- and 5-year survival rate for men and women with BC was comparable (Fig. 5D). Lymph node metastasis and the number of the lymph nodes affected as well as the number of distant metastases were also key factors in 3- and 5-year survival outcomes.

3.4. Validation of the nomogram's predictive accuracy

In the internal validation cohort, the c-index of the prognostic nomogram for overall survival was 0.7917, which was slightly higher compared to the training cohort (0.7916), suggesting high discrimination ability of the model. The c-index of the external validation cohort was 0.724 (95% CI, 0.66–0.79). The AUCs, an

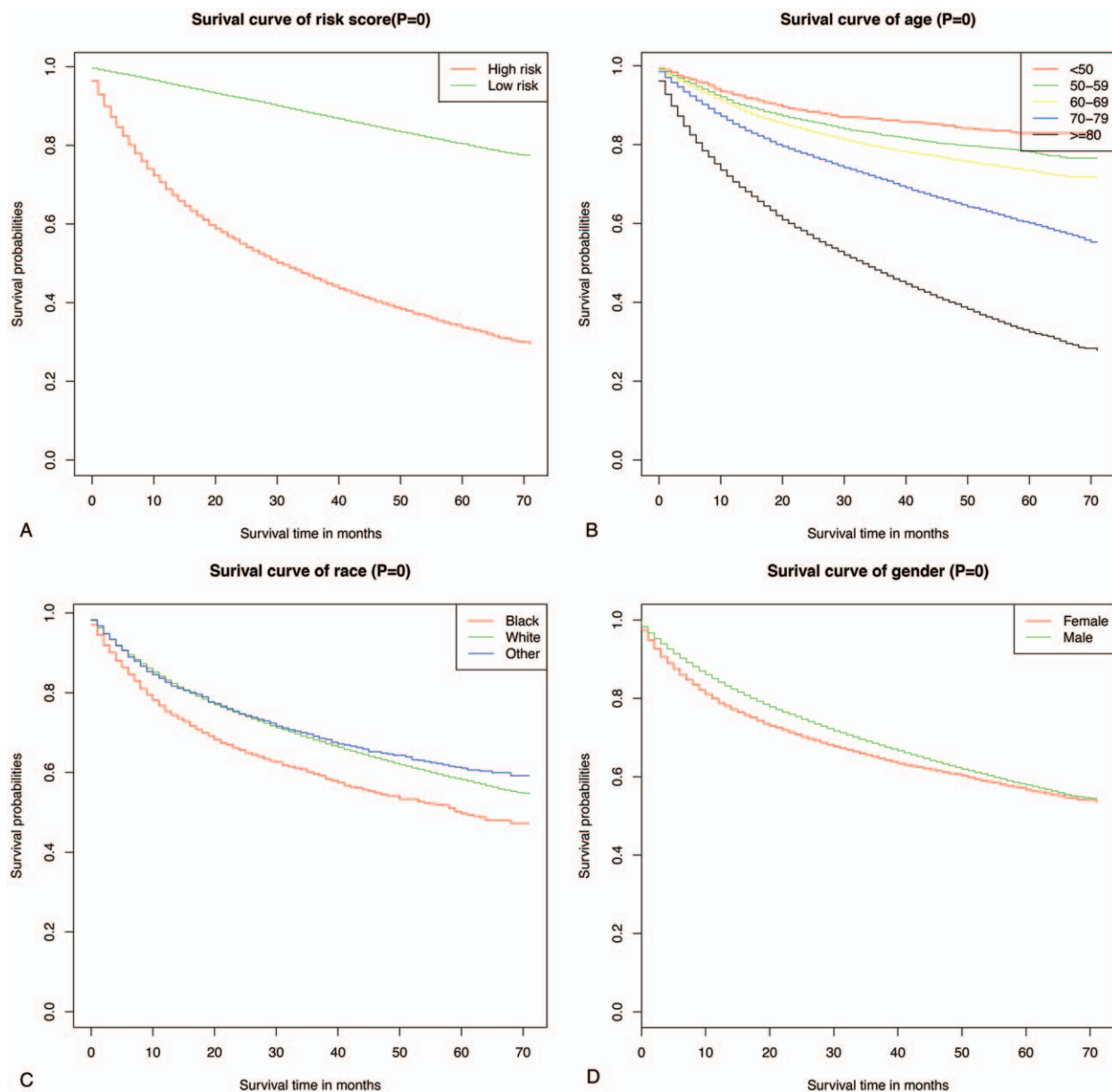


Figure 5. Predicted probability of overall survival by risk (A), age (B), race (C), sex (D) shown using Kaplan-Meier curve.

indicator of a model's discrimination ability, were 0.815 and 0.803 for the 3- and 5-year survival, respectively, in the internal validation cohort (Fig. 6A, B). The calibration plots showed excellent agreement between the internal and external validation cohorts (Fig. 7A–D). The c-index for the TNM-based model was superior to that for the AJCC-TNM classification-based model Whether it's the SEER training cohort (TNM-based model: 0.7916, 95% CI, 0.79–0.80 vs AJCC-TNM:0.739, 95% CI, 0.73–0.74) (see Table 1, Supplemental Content, <http://links.lww.com/MD/D309>, which illustrates the c-index for the TNM-based model in SEER) (see Table 2, <http://links.lww.com/MD/D309>, Supplemental Content, which illustrates the c-index for the AJCC-TNM classification-based model in SEER) or in the TCGA cohort (TNM-based model: 0.724, 95% CI, 0.66–0.79 vs AJCC-TNM: 0.689, 95% CI, 0.61–0.77) (see Table 3, <http://links.lww.com/MD/D309>, Supplemental Content, which illustrates the

c-index for the TNM-based model in TCGA cohort)(see Table 4, <http://links.lww.com/MD/D309>, Supplemental Content, which illustrates the c-index for the AJCC-TNM classification-based model in TCGA cohort).

4. Discussion

Urothelial carcinoma of the urinary bladder is a heterogeneous disease with multiple possible treatment modalities and a wide spectrum of clinical outcomes. Nomograms are considered a reliable graphical calculation model that combines all risk factors for tumor occurrence and have been used to predict individual risks of particular events.^[21,22] A tool that accurately evaluates the likelihood of metastatic progression, cancer-specific complications and mortality as well as long-term quality of life is important in patient counseling and decision-making.

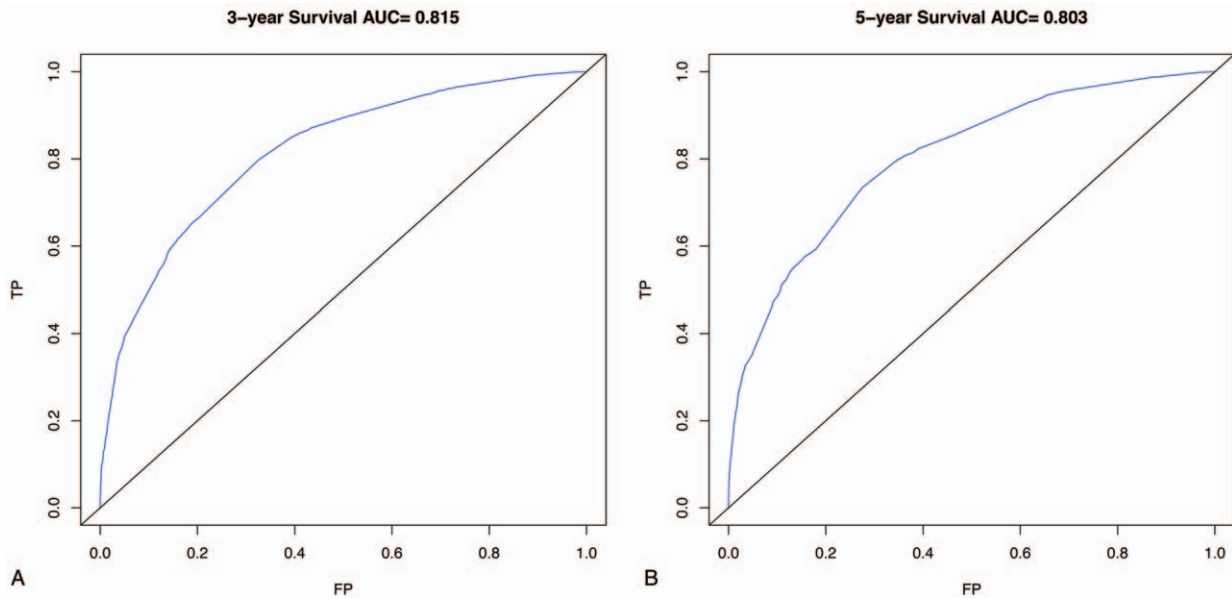


Figure 6. Area under the curves to predict overall survival at 3 years (A) and 5 years (B) using the internal validation cohort.

The model presented in this study was based on more than 60,000 patients from the SEER database and 130 patients in TCGA database. In this retrospective study, we evaluated the clinicopathological parameters of BC and independent prognostic factors that might have affected survival of patients included in the SEER database. We showed that age, sex, race and stage_T1, stage_T2a, stage_T2b, stage_T3a, stage-Ta, stage_Tis, stage_N, stage_M were independent prognostic factors. In addition, in multivariable analyses, we demonstrated that older age as well as more advanced T and N stages were independently associated with lower overall survival in BC.

Given these independent prognostic factors, we constructed a novel nomogram that combined TNM staging with some clinical parameters. The nomogram illustrated that age, the T stage, and M stage are the most significant contributors to prognosis, while gender and the N stage showed limited impact on outcomes. In addition, over the age of 60 years, for every 10 years of age, the risk of mortality increased multifold. The multivariate analysis revealed that age was the main determinant of a high-risk prognosis, followed by distant and lymph node metastases, and tumor invasion of the abdominal and pelvic wall. Nevertheless, age remains the most significant risk factor for BC mortality, as most BC cases occur among people over 60 years old. Other risk factors, such as smoking and exposure to some industrial chemicals, which have been shown to increase BC risk, should be accounted for in patient assessment. Our nomogram revealed an unexpected finding; in this study, some of the stage_T (T3b, T4a, T4b) sub-categories were not independent prognostic factors.

The TNM staging systems are commonly used for predicting patient prognosis. However, even in patients with the same stage of BC, there might be significant differences in prognosis and survival. In the present study, we developed a pictogram that predicts the overall survival. We observed that the c-index for the TNM-based model was superior to that for the AJCC-TNM classification (SEER training cohort: 0.7916 vs 0.739, $P < .05$; TCGA cohort: 0.724 vs 0.69, $P < .05$). The findings presented in this study demonstrate the differences in the clinical value of the

distinct risk evaluation systems, TNM- and AJCC-TNM-based systems, in estimating overall survival in BC patients. Moreover, the nomogram based on the TNM staging system was more effective in predicting patients' survival.

Meanwhile, the c-index for the TNM-based model in the validation cohort was slightly higher than for the internal training cohort based on high level score (training cohort 0.7916 vs internal validation cohort 0.7917), indicating that the nomogram we constructed was robust and accurate. The c-index for the external cohort was 0.724 (95% CI, 0.66–0.79). Although the c-index for the external validation cohort was not higher than the c-index for the training cohort (0.724), it was sufficient to confirm the stability and reliability of the prognostic model.

Calibration plots demonstrated excellent agreement between the nomogram-predicted and observed survival, which confirmed the validity and reliability of the novel nomogram. In addition, the survival curves showed that although the median survival time was 38 months, most patients did not survive beyond that point.

Regarding specific TNM stages, we found that Ta, Tis, and T1 stages were significant predictors of the overall survival compared to the other T-staging indicators. In clinical practice, BC at any of these three stages is referred to as non-muscular invasive bladder cancer (NMIBC). This suggests that NMIBC might be an independent prognostic factor of BC survival.

In addition, we divided our sample into risk levels based on median risk values derived from the KM curve analysis. Such fine-level grouping can support clinicians in assessing distinct prognoses for seemingly similar patients.

To the best of the authors' knowledge, this is the first nomogram derived based on a large dataset. The data used originated from the SEER database, ensuring the validity and reliability of our conclusions, as well as the internal and external validity of the nomogram, which offers an improvement in predictive accuracy over previously established models. To verify the value and prevent over-fitting of the present model, it was necessary to verify the novel nomogram.^[23–25] We have verified

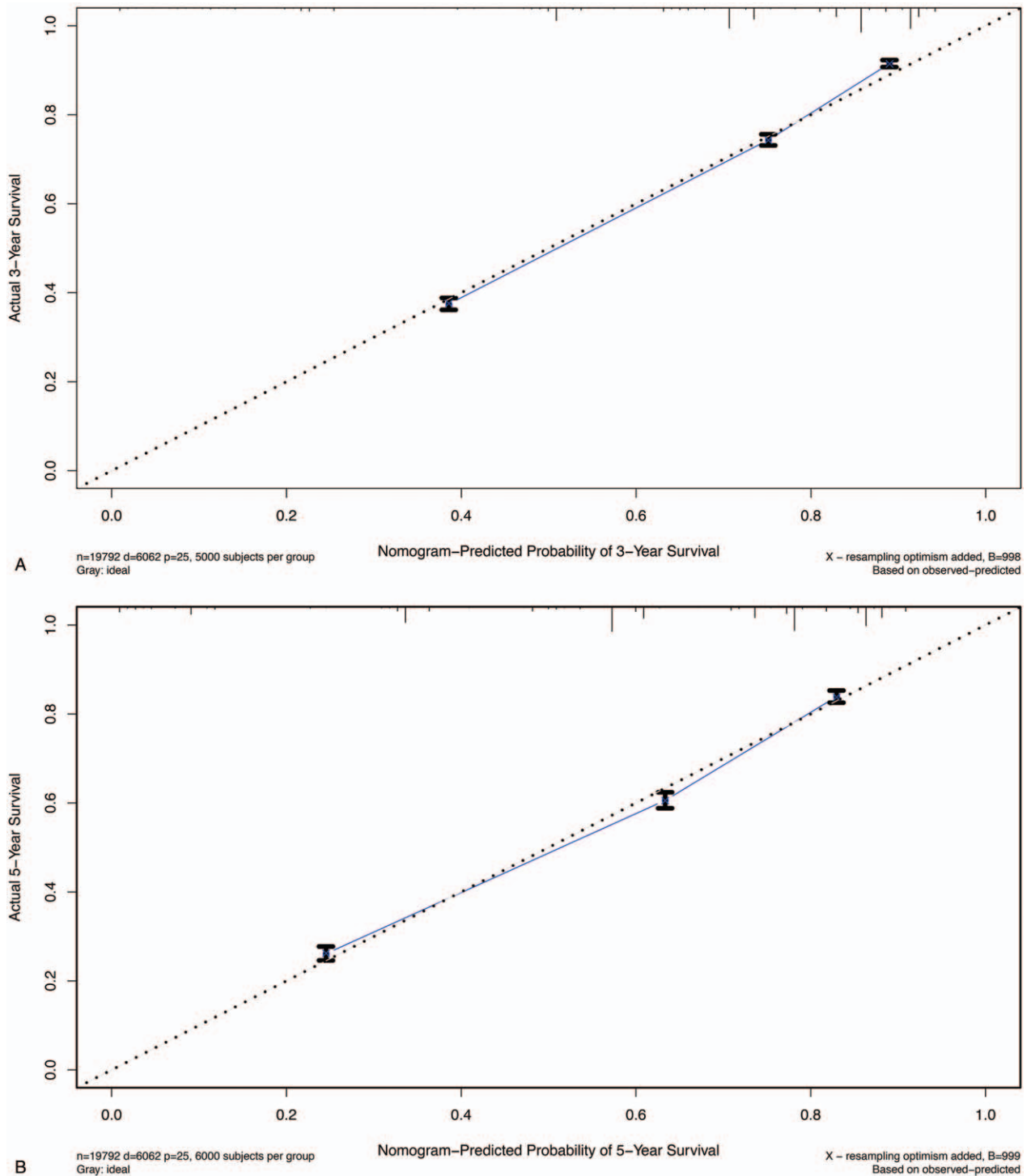


Figure 7. The calibration curves for predicting patient survival at (A) 3-year and (B) 5-year in internal validation cohort, and at (C) 3-year and (D) 5-year in external cohort. Nomogram-predicted survival is plotted on the x-axis, and the actual survival is plotted on the y-axis.

the predictive value of the model by using internal and external validation cohorts.

This study used the SEER and TCGA datasets; several similarities and differences between the TCGA and SEER data were observed. More than half of the patients with BC in the SEER database (60.2%) were over 70 years old, while there were slightly fewer patients in this age group in the TCGA database

(49.9%). Men were three times as likely to have BC compared to women (SEER F:M ratio: 23.4% vs 76.5%; TCGA F:M ratio: 20.7% vs 79.2%). As previously reported, the incidence of bladder cancer in men is three to four times higher than in women. Seven percent of all new cancer cases were men, but only women include 2% of new cases of the cancer.^[26] This is consistent with our findings. Most patients included in the SEER

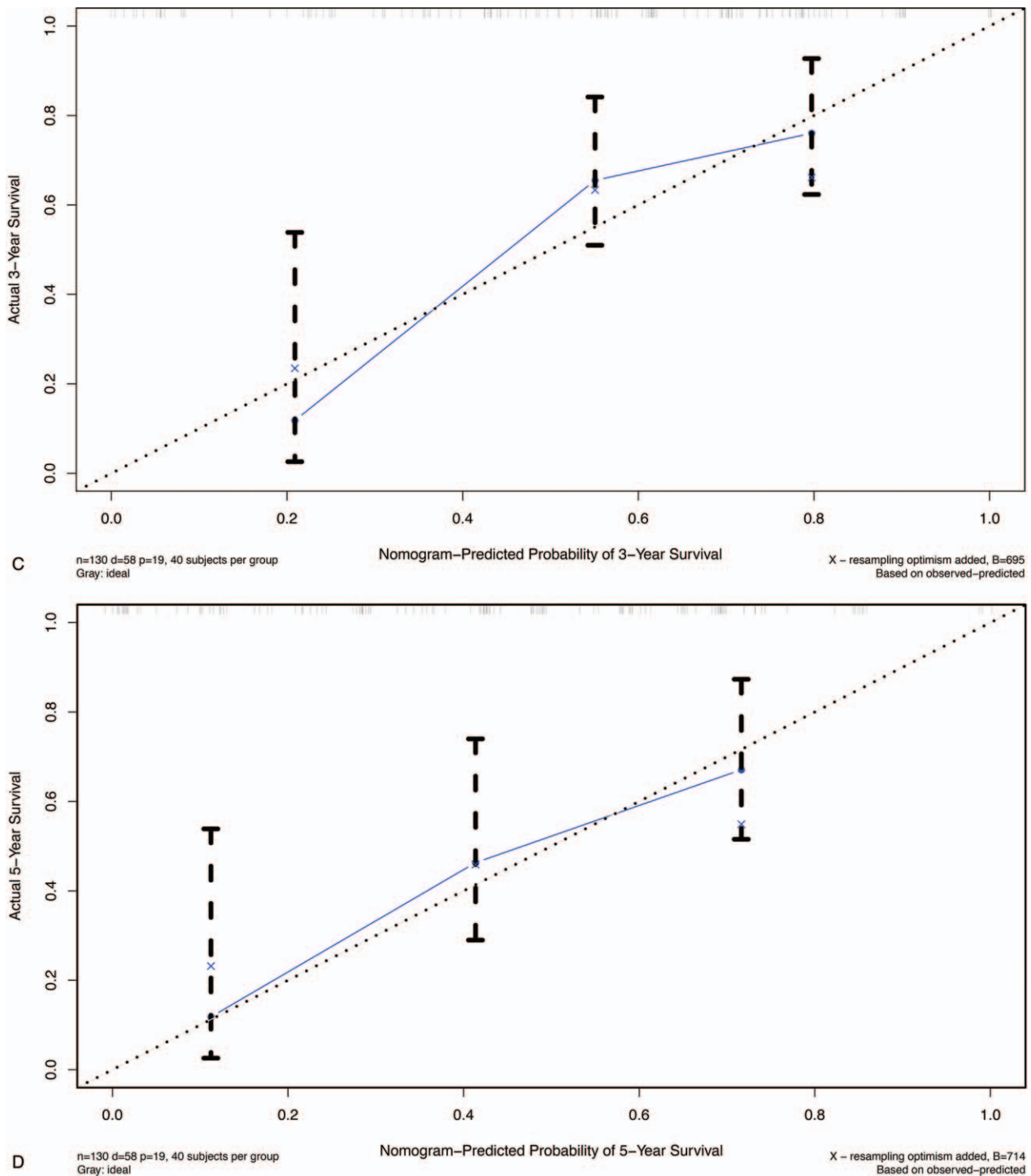


Figure 7. (Continued).

database were BC stage_Ta (45.4%). In contrast, in the TCGA database, most patients had BC stage_T3b (29.2%). However, T-stage classification was the strongest predictor of survival in our study, as reported by a cohort study based on the international Cancer Database.^[27] Given that this was retrospective analysis, the time coverage of the SEER database might have resulted in selection bias.

We used the TCGA database for external validation, as an independent validation cohort is required to confirm the validity

and reliability of a nomogram. The c-index of the external cohort was 0.724 (95% CI, 0.66–0.79), while the AUCs, another indicator of discrimination ability, were 0.815 and 0.803, for the 3- and 5-year survival, respectively. Moreover, calibration plots showed excellent agreement with the external validation cohort. Using external validation data provides additional evidence of the reliability, accuracy, and validity of our model. This notwithstanding, the TCGA database does not contain details on clinical diversity factors, such as tumor metastasis sites, tumor size, or

surgical methods used during patient treatment. As we were not able to account for these potentially relevant factors, our model might be missing potentially relevant variables that could further improve its predictive value. The Surveillance, Epidemiology, and End Results (SEER) database has been chosen in the current analysis given its high quality as well as large suitable cohort sizes. However, the SEER database is a retrospectively rather than a prospectively maintained resource, which might have biased our analysis or affected it in ways that we were not able to adjust for. At the time of writing, there are prospective, randomized studies underway, which aim to verify the validity and reliability of models predicting overall cancer survival. These studies use sophisticated data modeling, accounting for biochemical and immunological factors, such as tumor markers and genetic variants. However, until these results become available, no additional prognostic factors can be included in the nomogram.

It is worth noting that our model is based on the SEER database, which includes different races, further supporting the potential application of our nomogram to international populations. One of the key advantages of the proposed nomogram is the involvement of parameters that are easy to access and assess, meaning the model can be used in clinical practice without the burden of costs associated with complex tests, for example, tumor markers. Considering the size of the study population, these potential classification factors are unlikely to affect our conclusions. However, this nomogram was based on a retrospective study design; prospective clinical trials are needed to further validate this model.

Despite these limitations, the models presented in this study might be suitable for clinical use, supporting individualized assessment of expected survival in BC patients. They might also be used as a layered tool for clinical research, and evidence for development of interventions aimed at improving the overall survival.

Author contributions

Conceptualization: Ming-En Lin, Xue-Jun He

Data curation: Ying-Kai Hong

Funding acquisition: Ming-En Lin.

Methodology: Ye Zhang, Dong-Wu Zhuang.

Project administration: Ming-en Lin.

Software: Ye Zhang.

Visualization: Dong-Wu Zhuang, Ye Zhang.

Writing – original draft: Ye Zhang.

Writing – review & editing: Ye Zhang.

References

- [1] Antoni S, Ferlay J, Soerjomataram I, et al. Bladder cancer incidence and mortality: a global overview and recent trends. *Eur Urol* 2016;71:96–108.
- [2] Siegel RL, Miller KD, Jemal A. Cancer statistics. *CA: Cancer J Clin* 2018;68:7–30.
- [3] Kamat AM, Hahn NM, Efstathiou JA, et al. Bladder cancer. *Lancet* 2016;388:2796–810.
- [4] Zuiverloon TCM, de Jong FC, Theodorescu D. Clinical decision making in surveillance of non-muscle-invasive bladder cancer: the evolving roles of urinary cytology and molecular markers. *Oncology* 2017;31:855–62.
- [5] Pietzak EJ, Bagrodia A, Cha EK, et al. Next-generation sequencing of nonmuscle invasive bladder cancer reveals potential biomarkers and rational therapeutic targets. *Eur Urol* 2017;72:952–9.
- [6] Zhan Y, Du L, Wang L, et al. Expression signatures of exosomal long non-coding RNAs in urine serve as novel non-invasive biomarkers for diagnosis and recurrence prediction of bladder cancer. *Mol Cancer* 2018;17:142.
- [7] Bladder cancer: diagnosis and management of bladder cancer: (NICE (2015) Bladder cancer: diagnosis and management of bladder cancer. *BJU Int* 2017;120:755–65.
- [8] Amin MB, Greene FL, Edge SB, et al. The Eighth Edition AJCC Cancer Staging Manual: continuing to build a bridge from a population-based to a more “personalized” approach to cancer staging. *CA Cancer J Clin* 2017;67:93–9.
- [9] Validation of the Eighth AJCC new substages for bladder cancer among different staging contexts. *Clin Genitourin Cancer* 2017;15:e1095–106.
- [10] Wang X, Mao M, He Z, et al. Development and validation of a prognostic Nomogram in AFP-negative hepatocellular carcinoma. *Int J Biol Sci* 2019;15:221–8.
- [11] Fakhry C, Zhang Q, Nguyen-Tân PF, et al. Development and validation of nomograms predictive of overall and progression-free survival in patients with oropharyngeal cancer. *J Clin Oncol* 2017;35:4057–65.
- [12] Huang L, Balavarca Y, van der Geest L, et al. Development and validation of a prognostic model to predict the prognosis of patients who underwent chemotherapy and resection of pancreatic adenocarcinoma: a large international population-based cohort study. *BMC Med* 2019;17:66.
- [13] Cao J, Yuan P, Wang L, et al. Clinical nomogram for predicting survival of esophageal cancer patients after esophagectomy. *Sci Rep* 2016;6:26684.
- [14] Jiang S, Zhao R, Li Y, et al. Prognosis and nomogram for predicting postoperative survival of duodenal adenocarcinoma: a retrospective study in China and the SEER database. *Sci Rep* 2018;8:7940.
- [15] Cambier S, Sylvester RJ, Collette L, et al. EORTC nomograms and risk groups for predicting recurrence, progression, and disease-specific and overall survival in non-muscle-invasive stage Ta-T1 urothelial bladder cancer patients treated with 1-3 years of maintenance Bacillus Calmette-Guérin. *Eur Urol* 2016;69:60–9.
- [16] D Andrea D, Abufaraj M, Susani M, et al. Accurate prediction of progression to muscle-invasive disease in patients with pT1G3 bladder cancer: A clinical decision-making tool. *Urol Oncol* 2018;36:239.
- [17] Riestler M, Taylor JM, Feifer A, et al. Combination of a novel gene expression signature with a clinical nomogram improves the prediction of survival in high-risk bladder cancer. *Clin Cancer Res* 2012;18:1323–33.
- [18] Liu Y, Liu Z, Fu Q, et al. Galectin-9 as a prognostic and predictive biomarker in bladder urothelial carcinoma. *Urol Oncol* 2017;35:349–55.
- [19] Shariat SF, Karakiewicz PI, Ashfaq R, et al. Multiple biomarkers improve prediction of bladder cancer recurrence and mortality in patients undergoing cystectomy. *Cancer* 2008;112:315–25.
- [20] Peng D, Zhang CJ, Gong YQ, et al. Prognostic significance of HALP (hemoglobin, albumin, lymphocyte and platelet) in patients with bladder cancer after radical cystectomy. *Sci Rep* 2018;8:794.
- [21] Kluth LA, Black PC, Bochner BH, et al. Prognostic and prediction tools in bladder cancer: a comprehensive review of the literature. *Eur Urol* 2015;68:238–53.
- [22] Balachandran VP, Gonen M, Smith JJ, et al. Nomograms in oncology: more than meets the eye. *Lancet Oncol* 2015;16:e173–80.
- [23] Liang W, Zhang L, Jiang G, et al. Development and validation of a nomogram for predicting survival in patients with resected non-small-cell lung cancer. *J Clin Oncol* 2015;33:861–9.
- [24] Wei JH, Feng ZH, Cao Y, et al. Predictive value of single-nucleotide polymorphism signature for recurrence in localised renal cell carcinoma: a retrospective analysis and multicentre validation study. *Lancet Oncol* 2019;20:591–600.
- [25] Fang C, Wang W, Feng X, et al. Nomogram individually predicts the overall survival of patients with gastroenteropancreatic neuroendocrine neoplasms. *Br J Cancer* 2017;117:1544–50.
- [26] Ahmadi M, Ranjbaran H, Amiri MM, et al. Epidemiologic and socioeconomic status of bladder cancer in mazandaran province, northern Iran. *Asian Pac J Cancer Prev* 2012;13:5053–6.
- [27] Wakai K, Utsumi T, Yoneda K, et al. Development and external validation of a nomogram to predict high-grade papillary bladder cancer before first-time transurethral resection of the bladder tumor. *Int J Clin Oncol* 2018;23:957–64.