BMJ Open

Development and validation of nomograms to predict survival in patients with invasive micropapillary carcinoma of the breast

Yixin Cheng ,[1,2] Pengkun Zhang,[1,2] Yulin Huang,[1,2] Zhihui Zhang,[1,2] Ru Tang,[1,2] Feng Chi,[1,2] Jia-Yuan Sun,[1,2] Zhenyu He[1,2]

To cite: Cheng Y, Zhang P, Huang Y, *et al*. Development and validation of nomograms to predict survival in patients with invasive micropapillary carcinoma of the breast. *BMJ Open* 2023;**13**:e065312. doi:10.1136/bmjopen-2022-065312

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2022-065312).

YC, PZ and YH contributed equally.

Received 01 June 2022
Accepted 08 February 2023

[1]Department of Radiation Oncology, Sun Yat-Sen University Cancer Center, Guangzhou, Guangdong, China
[2]State Key Laboratory of Oncology in South China, Sun Yat-Sen University, Guangzhou, Guangdong, China

Correspondence to
Dr Zhenyu He;
hezhy@sysucc.org.cn

ABSTRACT

Objectives The present study aimed to develop and validate nomograms to predict the survival of patients with breast invasive micropapillary carcinoma (IMPC) to aid objective decision-making.

Design Prognostic factors were identified using Cox proportional hazards regression analyses and used to construct nomograms to predict overall survival (OS) and breast cancer-specific survival (BCSS) at 3 and 5 years. Kaplan-Meier analysis, calibration curves, the area under the curve (AUC) and the concordance index (C-index) evaluated the nomograms' performance. Decision curve analysis (DCA), integrated discrimination improvement (IDI) and net reclassification improvement (NRI) were used to compare the nomograms with the American Joint Committee on Cancer (AJCC) staging system.

Setting Patient data were collected from the Surveillance, Epidemiology, and End Results (SEER) database. This database holds data related to the incidence of cancer acquired from 18 population-based cancer registries in the US.

Participants We ruled out 1893 patients and allowed the incorporation of 1340 patients into the present study.

Results The C-index of the AJCC8 stage was lower than that of the OS nomogram (0.670 vs 0.766) and the OS nomograms had higher AUCs than the AJCC8 stage (3 years: 0.839 vs 0.735, 5 years: 0.787 vs 0.658). On calibration plots, the predicted and actual outcomes agreed well, and DCA revealed that the nomograms had better clinical utility compared with the conventional prognosis tool. In the training cohort, the NRI for OS was 0.227, and for BCSS was 0.182, while the IDI for OS was 0.070, and for BCSS was 0.078 (both p<0.001), confirming its accuracy. The Kaplan-Meier curves for nomogram-based risk stratification showed significant differences (p<0.001).

Conclusions The nomograms showed excellent discrimination and clinical utility to predict OS and BCSS at 3 and 5 years, and could identify high-risk patients, thus providing IMPC patients with personalised treatment strategies.

STRENGTHS AND LIMITATIONS OF THIS STUDY

⇒ Prognostic factors were identified using Cox proportional hazards regression analyses and used to construct nomograms to predict the 3-year and 5-year prognosis of patients with invasive micropapillary carcinoma (IMPC).
⇒ To compare the accuracy of the nomograms with that of American Joint Committee on Cancer 8 staging, the integrated discrimination improvement, net reclassification improvement and decision curve analysis were calculated.
⇒ By calculating the score of all the variables, we divided patients with IMPC into low-risk and high-risk groups.
⇒ Although this was a large-sample study, it was retrospective in nature, which might have generated confounding and it is difficult to conclude casual basis of any association.

INTRODUCTION

Fisher *et al* first described invasive micropapillary carcinoma (IMPC) of the breast in 1980[1] and this rare breast cancer variant was further characterised by Siriaunkgul and Tavassoli 13 years later.[2] IMPC accounts for approximately 3%–6% of all invasive breast cancers[3] and was classified by the WHO as an independent subtype of breast cancer in 2003.[4] IMPC has an aggressive nature and studies have revealed that IMPC has a high frequency of lymphovascular invasion (LVI), regional lymph nodal metastasis and local recurrence.[5–7]

Currently, to predict prognosis, clinicians commonly use the American Joint Committee on Cancer (AJCC) staging system. However, because of the rarity and aggressive behaviour of IMPC, prognosis prediction based only on the AJCC stage is insufficient to meet the increasing need for personalised medicine.[8] In addition to the factors included in the tumor–node–metastasis classification, evidence indicates that other elements, such as demographic characteristics and treatment strategies, impact the prognosis of patients with IMPC.[9–12] Therefore, it is important to

consider and identify factors that can serve as prognostic factors to build models to predict the survival outcomes of patients with IMPC accurately. Besides, a risk classification system incorporating the variables above is important because it is the premise of optimal patients' treatment.

Nomograms, as reliable and practical evaluation tools, are being used increasingly in clinical oncology.[13 14] They can quantitatively predict the prognosis of a given patient using multivariate analysis-derived prognostic factors and provide visualised prediction results. Given the many clinicopathological characteristics that might influence the prognosis and annually increasing incidence rate of IMPC,[3] there is an urgent need to establish a credible and comprehensive model.

The present study aimed to use data available from the Surveillance, Epidemiology, and End Results (SEER) database to build nomograms to predict breast cancer-specific survival (BCSS) and overall survival (OS) and to identify patients with different risk levels. We also aimed to test the prognostic and clinical value of the nomograms in comparison with the AJCC8 staging system.

MATERIALS AND METHODS

Patient and public involvement

No patient involved.

Selection of patients and processing of data

Patient data were collected from the SEER database. This database holds data related to the incidence of cancer acquired from 18 population-based cancer registries that represent about 30% of the population of the USA. We acquired permission in November 2020 to analyse the study data (Username: 16366-Nov2020).

The following specific inclusion criteria were used: (1) histology ICD-O-3 was restricted to IMPC (8507/3), (2) site record ICD-O-3/WHO 2008 was restricted to breast and (3) year of diagnosis from 2010 to 2015. The following exclusion criteria were used: (1) patients lacking a definite pathological diagnosis, (2) patients whose survival data was incomplete or inaccessible, (3) detailed information was lacking for sex, race or age, (4) patients without grade classification, AJCC stage or without metastatic sites. Using these criteria ruled out 1893 patients and allowed the incorporation of 1340 patients into the present study.[15]

Covariates and endpoint

The variables assessed included marital status, race, age, location, laterality, grade, metastatic sites, AJCC stage, breast subtype, surgery, chemotherapy, radiotherapy and follow-up information. We used the four-grade system as in SEER Instructions for Coding Grade. In this study, OS and BCSS were the endpoints. OS was defined as the period from diagnosis to death from any cause or the date of the last follow-up for patients that remain alive. BCSS was defined as the period from diagnosis to death that was attributed to breast cancer or date of the last follow-up for patients that remain alive.

Nomogram development and statistical analyses

The patients were divided randomly into a training cohort (n=937) and a validation cohort (n=403) at a 7:3 ratio. Categorical variables, displayed as proportions and frequencies, were compared employing Fisher's exact test or the χ^2 test. Continuous variable such as time were compared using Mann-Whitney U test. Univariate analysis identified potential prognostic variables that significantly ($p<0.05$) affected OS and BCSS, which were then subjected to multivariate analysis. Integration of these identified predictors allowed the construction of nomograms to predict the 3-year and 5-year prognosis of patients with IMPC.

The nomograms were subjected to internal and external validation. The nomograms' discriminative abilities were assessed using the concordance index (C-index) and receiver operating characteristic curves. We plotted calibration curves for the comparison of actual patient survival with nomogram-predicted survival. We used bootstrapping with 1000 resampling events to evaluate discrimination and calibration. In addition, to compare the accuracy of the nomograms with that of AJCC8 staging, the integrated discrimination improvement (IDI) and net reclassification improvement (NRI) were calculated. The clinical efficacy of the nomograms and AJCC8 staging was compared using decision curve analysis (DCA).

Then, we calculated the score of all the variables. An aggregate score from the nomograms was assigned to each patient. Then, the sum score was used to divide patients with IMPC into low-risk and high-risk groups. To evaluate the prognostic differences between the two risk groups, the Kaplan-Meier method was used for survival analysis employing the log-rank test.

The R software V.4.1.1 was used for all statistical analyses and to determine the optimal cut-off values. Statistical significance was accepted at a p value<0.05.

RESULTS

Characteristics of the included patients

The selection criteria identified 1340 eligible patients with IMPC from the SEER database, which were divided using the random split-sample method into a training cohort (n=937) and a validation cohort (n=403). At diagnosis, the patient's median age was 62 years old and they had a median follow-up time of 56 months (IQR, 43–74 months).[15] No statistically significant differences were found among the patients' clinicopathological and demographic and features (table 1).

Independent prognostic factors for OS and BCSS

In the training cohort, seven independent prognostic factors were identified using univariate Cox regression analysis followed by multivariate analysis for OS:

Table 1 Patient demographics and pathological characteristics[15]

	Level	Overall	Training	Validation	P value
n		1340	937	403	
Age (%)	<60	579 (43.21)	415 (44.29)	164 (40.69)	0.2467
	≥60	761 (56.79)	522 (55.71)	239 (59.31)	
Sex (%)	Female	1316 (98.21)	922 (98.40)	394 (97.77)	0.5647
	Male	24 (1.79)	15 (1.60)	9 (2.23)	
Race (%)	Black	172 (12.84)	120 (12.81)	52 (12.90)	0.989
	White	1024 (76.42)	717 (76.52)	307 (76.18)	
	Others*	144 (10.75)	100 (10.67)	44 (10.92)	
Marital (%)	Married	712 (53.13)	499 (53.26)	213 (52.85)	0.9732
	Unmarried	563 (42.01)	392 (41.84)	171 (42.43)	
	Unknown	65 (4.85)	46 (4.91)	19 (4.71)	
Laterality (%)	Left	655 (48.88)	452 (48.24)	203 (50.37)	0.5113
	Right	685 (51.12)	485 (51.76)	200 (49.63)	
Location (%)	Others/NOS	952 (71.04)	677 (72.25)	275 (68.24)	0.1556
	Upper-outer quadrant	388 (28.96)	260 (27.75)	128 (31.76)	
Grade (%)	I/II	862 (64.33)	590 (62.97)	272 (67.49)	0.1275
	III/IV	478 (35.67)	347 (37.03)	131 (32.51)	
AJCC7th (%)	I	578 (43.13)	409 (43.65)	169 (41.94)	0.8985
	II	443 (33.06)	308 (32.87)	135 (33.50)	
	III	266 (19.85)	182 (19.42)	84 (20.84)	
	IV	53 (3.96)	38 (4.06)	15 (3.72)	
Bone (%)	No	1314 (98.06)	921 (98.29)	393 (97.52)	0.468
	Yes	26 (1.94)	16 (1.71)	10 (2.48)	
Brain (%)	No	1337 (99.78)	935 (99.79)	402 (99.75)	1
	Yes	3 (0.22)	2 (0.21)	1 (0.25)	
Liver (%)	No	1334 (99.55)	933 (99.57)	401 (99.50)	1
	Yes	6 (0.45)	4 (0.43)	2 (0.50)	
Lung (%)	No	1332 (99.40)	933 (99.57)	399 (99.01)	0.3976
	Yes	8 (0.60)	4 (0.43)	4 (0.99)	
Subtype (%)	HR−/HER2− (triple negative)	54 (4.03)	40 (4.27)	14 (3.47)	0.9028
	HR−/HER2+ (HER2 enriched)	60 (4.48)	43 (4.59)	17 (4.22)	
	HR+/HER2− (luminal A)	1001 (74.70)	697 (74.39)	304 (75.43)	
	HR+/HER2+ (luminal B)	225 (16.79)	157 (16.76)	68 (16.87)	
Surgery (%)	Breast-conserving surgery	650 (48.51)	449 (47.92)	201 (49.88)	0.8057
	Mastectomy	625 (46.64)	442 (47.17)	183 (45.41)	
	No/unknown	65 (4.85)	46 (4.91)	19 (4.71)	
Radiotherapy (%)	No/unknown	597 (44.55)	417 (44.50)	180 (44.67)	1
	Yes	743 (55.45)	520 (55.50)	223 (55.33)	
Chemotherapy (%)	No/unknown	663 (49.48)	462 (49.31)	201 (49.88)	0.8952
	Yes	677 (50.52)	475 (50.69)	202 (50.12)	

Continued

Table 1 Continued

Level	Overall	Training	Validation	P value
Time (median (IQR))	56.000 (43.000, 74.000)	55.000 (42.000, 74.000)	56.000 (43.500, 76.000)	0.5388

Unmarried, separated, divorced, widowed, single.
HR,strogen and Progesterone Receptors HER2 , human epidermal growth factor rec
*Others, American Indian/AK Native, Asian/Pacific Islander and unknown.
†Others, upper-inner, lower-outer, lower-inner, overlapping lesion of breast, axillary tail of the breast, nipple and central portion.
AJCC, The American Joint Committee for Cancer; NOS, not otherwise specified.

radiotherapy, chemotherapy, surgery, subtype, AJCC stage, marital status and age (table 2).

In addition, seven independent prognostic factors for BCSS (marital status, grade, AJCC stage, race, subtype, surgery and radiotherapy) were identified in the training cohorts (table 3).

Construction of the nomograms and their ability to stratify patient risk

Using the selected predictors from the training cohort, all the independent variables were combined to construct nomograms to predict OS and BCSS at 3 and 5 years. Figure 1 shows the 3-year and 5-year OS prognostic nomogram and figure 2 shows the 3-year and 5-year BCSS prognostic nomogram.

The probability of OS and BCSS at 3 and 5 years could be predicted by summing the total nomogram scores for each patient. Worse prognosis was indicated by a higher score. We calculated the best cut-off values using the R package 'survival'. Based on the OS nomogram-derived cut-off value, patients with IMPC could be classified into low-risk (score≤100) and high-risk (score>100) groups. Similarly, for BCSS, patients were classified as two subgroups (190≤ and >190). Subsequently, according to the Kaplan-Meier survival curve analysis, we observed an obvious grading ability in the new prognostic nomograms (p<0.0001) (figure 3). For both OS and BCSS in the validation cohort, the low-risk group had a favourable prognosis for both OS and BCSS (online supplemental figure 1).

Validation and calibration of the nomograms

The discriminative ability of the nomograms over different survival periods was assessed. Higher bootstrapped time-dependent C-indices for the prediction of OS and BCSS based on the nomograms were observed compared with derived from AJCC staging in both the training cohort (figure 4) and the validation cohort (online supplemental figure 2). For instance, the C-index for OS at 5 years from the nomogram (training group=0.766, validation group=0.794) exceeded that of AJCC8 staging (training set=0.670, validation set=0.732). Similarly, the C-index derived from the BCSS nomogram (training set=0.812, validation set=0.871) exceeded that of AJCC8 staging (training set=0.773, validation set=0.815). Furthermore, higher area under the curves (AUCs) were observed for the nomograms compared with those of AJCC staging in the training cohort (3-year AUC: 0.839 vs 0.735, 5-year AUC: 0.787 vs 0.658) and validation cohort (3-year AUC: 0.857 vs 0.773, 5-year AUC: 0.836 vs 0.712) for OS (figure 5). The results for BCSS are shown in online supplemental figure 3. These results suggested better discrimination by the nomograms than by AJCC8 staging. For 3-year and 5-year OS (figure 6) and BCSS (online supplemental figure 4), in both the training and validation cohorts, the nomograms displayed good agreement between the predicted and actual outcomes according to the calibration curves.

Moreover, we used the NRI and IDI to demonstrate the accuracy of our novel nomograms in comparison with AJCC8 staging. For 3-year and 5-year OS, the NRI in the training cohort was 0.197 (95% CI: 0.056 to 0.326) and 0.227 (95% CI: 0.123 to 0.328), respectively, and for 3-year and 5-year BCSS, the NRI was 0.086 (95% CI: 0.027 to 0.336) and 0.182 (95% CI: 0.021 to 0.326), respectively. For 3-year and 5-year OS, the IDI was 0.100 (p<0.001) and 0.070 (p<0.001), respectively, and or 3-year and 5-year BCSS, the IDI was 0.059 (p<0.001) and 0.078 (p<0.001) in the training cohort, respectively. Thus, the constructed nomograms displayed better accuracy to predict prognosis compared with AJCC8 staging.

Finally, DCA was used to compare the clinical utility of the nomograms with that of the traditional staging system. The 3-year and 5-year DCA curves showed that the nomograms have favourable clinical utilisation and benefits in the validation group compared with those AJCC staging (figure 7).

DISCUSSION

Among subtypes of breast cancer, IMPC is rare, only representing 0.2% of all breast cancers in the SEER database. Compared with, for example, invasive ductal carcinoma, IMPC is associated with more aggressive clinicopathological features. The distinct clinical features and small sample size of IMPC mean that there is no specific method to predict survival or a standard treatment strategy. Besides, it remains controversial whether IMPC correlates with worse survival.[16–20] Although the conventional AJCC staging system is a good tool to predict prognosis, it lacks important risk factors, which might reduce the accuracy of prediction. Thus, we determined that it was necessary to develop prognostic nomograms

Table 2 Univariate and multivariate Cox analysis of overall survival

	Univariate analysis		Multivariate analysis	
Characteristics	**HR (95% CI)**	**P value**	**HR (95% CI)**	**P value**
Age				
<60	Reference		Reference	
≥60	1.89 (1.32 to 2.71)	<0.001	1.98 (1.32 to 2.96)	<0.001
Sex				
Female	Reference			
Male	0.82 (0.2 to 3.32)	0.781		
Race				
Black	Reference		Reference	
White	0.64 (0.42 to 0.99)	0.044	0.78 (0.5 to 1.23)	0.291
Others*	0.4 (0.19 to 0.83)	0.014	0.52 (0.25 to 1.12)	0.094
Marital				
Married	Reference		Reference	
Unknown	1.36 (0.58 to 3.16)	0.48	1.32 (0.55 to 3.12)	0.534
Unmarried	2.34 (1.65 to 3.32)	<0.001	1.57 (1.08 to 2.28)	0.018
Location				
Others†	Reference			
Upper-outer quadrant	1.06 (0.73 to 1.52)	0.767		
Laterality				
Left	Reference			
Right	1.06 (0.76 to 1.48)	0.724		
Grade				
I/II	Reference		Reference	
III/IV	1.42 (1.02 to 1.98)	0.039	1.31 (0.91 to 1.89)	0.141
AJCC stage				
I	Reference		Reference	
II	1.29 (0.84 to 1.98)	0.243	1.55 (0.99 to 2.44)	0.056
III	2.16 (1.4 to 3.33)	0.001	5.05 (3.03 to 8.43)	<0.001
IV	6.81 (3.93 to 11.83)	<0.001	9.3 (4.75 to 18.21)	<0.001
Subtype				
HR–/HER2– (triple negative)	Reference		Reference	
HR–/HER2+ (HER2 enriched)	0.21 (0.08 to 0.58)	0.003	0.24 (0.09 to 0.69)	0.008
HR+/HER2– (luminal A)	0.27 (0.16 to 0.46)	<0.001	0.22 (0.12 to 0.39)	<0.001
HR+/HER2+ (luminal B)	0.3 (0.16 to 0.57)	<0.001	0.28 (0.14 to 0.55)	<0.001
Surgery				
Breast-conserving surgery	Reference		Reference	
Mastectomy	1.34 (0.93 to 1.92)	0.112	0.71 (0.47 to 1.09)	0.1147
No/unknown	5.82 (3.5 to 9.66)	<0.001	1.97 (1.07 to 3.63)	0.0305
Chemotherapy				
No	Reference		Reference	
Yes	0.55 (0.39 to 0.78)	0.001	0.37 (0.25 to 0.56)	<0.001
Radiotherapy				
No	Reference		Reference	
Yes	0.4 (0.29 to 0.57)	<0.001	0.48 (0.32 to 0.72)	<0.001

Unmarried, separated, divorced, widowed, single.
*Others, American Indian/AK Native, Asian/Pacific Islander and unknown.
†Others, upper-inner, lower-outer, lower-inner, overlapping lesion of the breast, axillary tail of breast, nipple,and central portion.
AJCC, The American Joint Committee for Cancer.;

Table 3 Univariate and multivariate Cox analysis of breast cancer-specific survival

	Univariate analysis		Multivariate analysis	
Characteristics	**HR (95% CI)**	**P value**	**HR (95% CI)**	**P value**
Age				
<60	Reference		Reference	
≥60	0.81 (0.52 to 1.28)	0.374		
Sex				
Female	Reference			
Male	0.78 (0.11 to 5.62)	0.806		
Race				
Black	Reference		Reference	
White	0.18 (0.05 to 0.61)	0.006	0.26 (0.08 to 0.9)	0.034
Others*	0.48 (0.28 to 0.82)	0.007	0.77 (0.44 to 1.34)	0.355
Marital				
Married	Reference		Reference	
Unknown	2.44 (1.01 to 5.89)	0.047	2.98 (1.22 to 7.3)	0.017
Unmarried	2.06 (1.27 to 3.33)	0.003	1.89 (1.16 to 3.09)	0.011
Location				
Others†	Reference			
Upper-outer quadrant	0.84 (0.49 to 1.43)	0.52		
Laterality				
Left	Reference			
Right	1.44 (0.91 to 2.28)	0.124		
Grade				
I/II	Reference		Reference	
III/IV	2.76 (1.74 to 4.4)	<0.001	2 (1.22 to 3.27)	0.006
AJCC stage				
I	Reference		Reference	
II	2.58 (1.24 to 5.35)	0.011	2.04 (0.97 to 4.29)	0.06
III	6.09 (3.03 to 12.23)	<0.001	6.36 (2.98 to 13.53)	<0.001
IV	21.72 (9.96 to 47.36)	<0.001	11.71 (4.67 to 29.35)	<0.001
Subtype				
HR–/HER2– (triple negative)	Reference		Reference	
HR–/HER2+ (HER2 enriched)	0.25 (0.08 to 0.79)	0.018	0.2 (0.06 to 0.67)	0.009
HR+/HER2– (luminal A)	0.21 (0.11 to 0.4)	<0.001	0.23 (0.11 to 0.47)	<0.001
HR+/HER2+ (luminal B)	0.18 (0.07 to 0.43)	<0.001	0.14 (0.06 to 0.35)	<0.001
Surgery				
Breast-conserving surgery	Reference		Reference	
Mastectomy	2.34 (1.36 to 4.02)	0.002	0.96 (0.52 to 1.77)	0.897
No/unknown	10.45 (5.23 to 20.89)	<0.001	2.44 (1.02 to 5.84)	0.046
Chemotherapy				
No	Reference		Reference	
Yes	0.92 (0.59 to 1.45)	0.735		<0.001
Radiotherapy				
No	Reference		Reference	
Yes	0.46 (0.29 to 0.74)	0.001	0.52 (0.31 to 0.87)	0.014

Unmarried, separated, divorced, widowed, single.
*Others, American Indian/AK Native, Asian/Pacific Islander and unknown.
†Others, upper-inner, lower-outer, lower-inner, overlapping lesion of the breast, axillary tail of the breast, nipple and central portion.
AJCC, The American Joint Committee for Cancer.

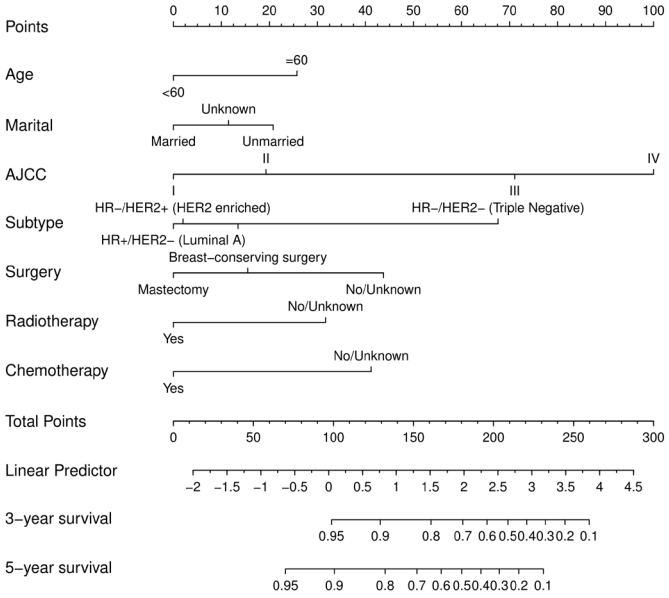

Figure 1 Nomogram to predict the overall survival of patients with invasive micropapillary breast cancer. AJCC, American Joint Committee on Cancer.

of OS and BCSS and risk grades that incorporate these factors for patients with IMPC of the breast. We believe our models will guide clinicians directly to identify high-risk patients and thereby design personalised therapeutic strategies.

In the present study, we used easily accessible clinicopathological factors and therapeutic schedules to construct nomograms that will be convenient for clinicians to use. Analysis of data from a large population-based database revealed seven risk factors for BCSS and OS. In our nomogram, radiotherapy, surgery, molecular subtypes, AJCC stage and marital status correlated significantly

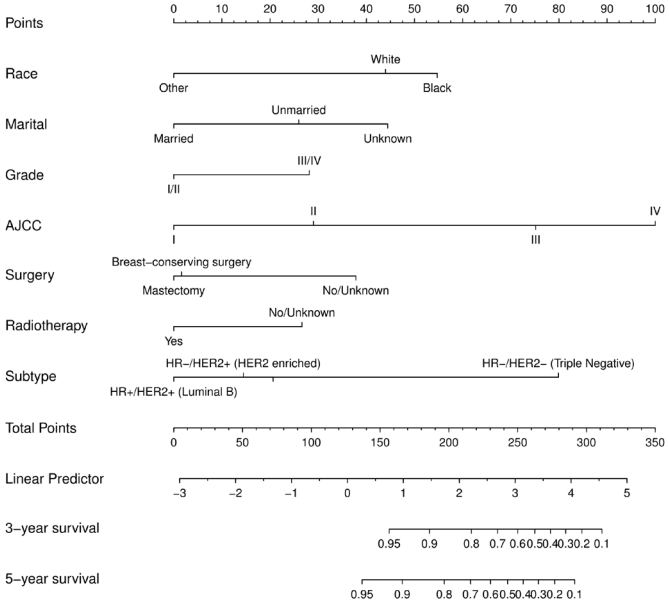

Figure 2 Nomogram to predict the breast cancer-specific survival of patients with invasive micropapillary breast cancer. AJCC, American Joint Committee on Cancer.

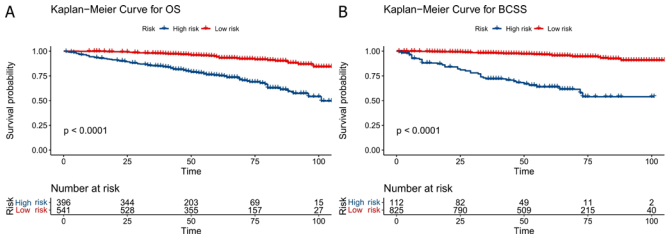

Figure 3 Analysis of patient survival post risk-stratification. (A) Results for overall survival (OS) in the training group. (B) Results for breast cancer-specific survival (BCSS) in the training group.

with both BCSS and OS, which agreed with the results of previous studies.[9 11 21–23] Among these factors, AJCC stage and molecular subtypes had an enormous impact on the nomograms. In agreement with the conclusion reported in some retrospective studies, black ethnicity was associated with poorer OS compared with other ethnicities.[10 24] In contrast to previous findings that HER2-enriched and triple-negative subtypes are related to poorer prognosis,[20] we found that the HER2-enriched subtype is related to a better prognosis than luminal B, which is HER2 negative. This might be attributed to the molecular-targeted HER2 therapy provided to patients with the HER2-enriched subtype.[25] Similar to previous SEER studies,[20 26] about half of the patients with IMPC underwent mastectomy, and the prognosis of these patients was slightly worse compared with those who received breast-conserving surgery. Meanwhile, a previous finding also demonstrated that the survival outcomes of patients with early-stage IMPC who underwent breast-conserving surgery were not inferior to those with mastectomy.[12] Moreover, we confirmed that patients with IMPC benefited significantly from radiotherapy in terms of both OS and BCSS, an issue that has long been controversial.[26 27] Additionally, consistent with a previous study, chemotherapy was inappropriate for use as a predictor for BCSS after adjustment for other factors.[28] Thus, further research is required on treatment regimens for patients with IMPC of the breast.

Based on the total scores of the nomograms, patients could be divided into high-risk or low-risk groups. Patients in the high-risk group had a statistically significantly poorer survival outcome compared with that of patients in the low-risk group. Consequently, the nomogram-based risk stratifications could provide clinicians with an accurate reference to distinguish those high-risk patients

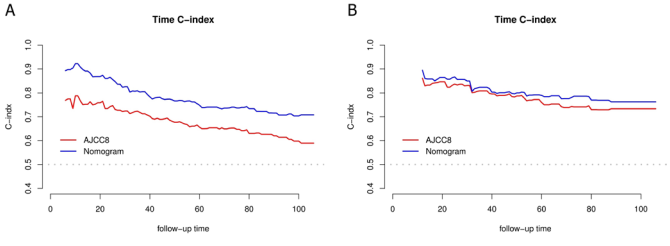

Figure 4 Time-dependent C-indices for overall survival (A) and breast cancer-specific survival (B) in the training set. AJCC, American Joint Committee on Cancer.

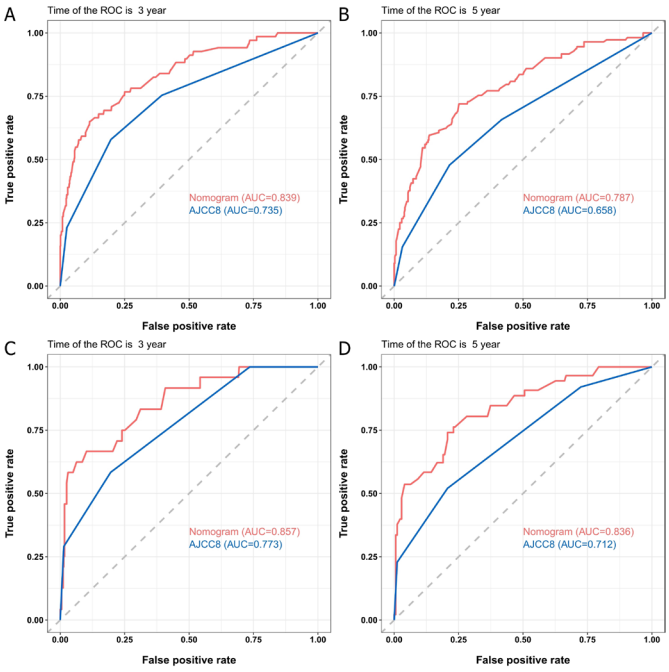

Figure 5 Comparison of survival prediction between AJCC TNM staging and the nomogram using time-dependent ROC curves. (A) Results for overall survival (OS) at 3 years the training group. (B) Results for OS at 5 years the training group. (C) Results for OS at 3 years the validation group. (D) Results for OS at 5 years the validation group. AJCC, The American Joint Committee for Cancer; ROC, receiver operating characteristic; TNM, tumor–node–metastasis.

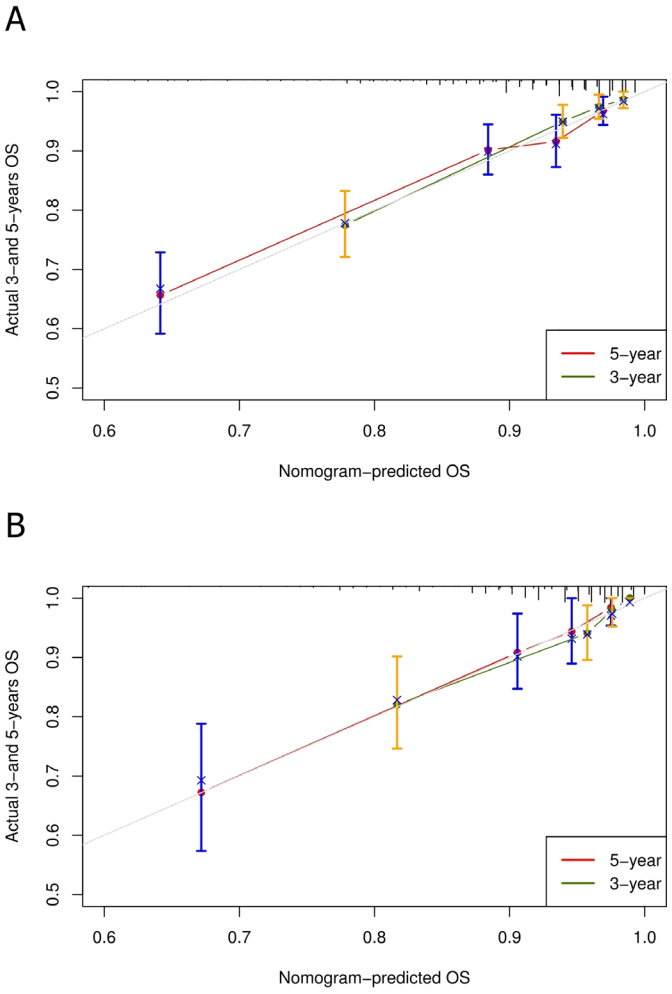

Figure 6 Calibration plots to predict survival at 3 and 5 years. (A) Results for overall survival (OS) in the training set. (B) Results for OS in the validation set.

that require a more active treatment strategy and could prevent overtreatment of low-risk patients.

Our nomograms displayed significantly higher time-dependent C-index and AUC values compared with those of the AJCC8 staging criteria, demonstrating superior discriminative power for predicting OS and BCSS. In addition, the actual survival and the nomogram-predicted survival agreed well according to the calibration plots, which indicated the reliability of the new models. Furthermore, the results of DCA proved that, compared with the traditional tool, the novel nomograms were better at predicting survival. The results of NRI and IDI also supported the view that our prediction models were effective and accurate.

In the present study, nomograms were developed to visualise the BCSS and OS for patients with IMPC of the breast based on a large dataset. We also compared the novel models with the traditional AJCC staging system, and the nomograms displayed a better prediction capacity and clinical utility than the AJCC staging system. Compared with a previous nomogram,[28] our nomograms have some improvements. First, we took molecular subtypes into consideration, including HER2 status, which has an important function in cancer prognosis and progression.[29] Then, we calculated the DCA of the nomograms. The wider range of threshold probabilities reflected a better clinical utility and benefits than that of the AJCC staging system. Therefore, clinicians would find the nomograms more beneficial for clinical management.

Nevertheless, this study had a number of limitations. First, although this was a large-sample study, it was retrospective in nature, which might have generated potential selection bias. Second, data from SEER database did not include certain specific information, such as LVI, BRCA1/2 mutation, or the Oncotype DX Recurrence Score. However, these factors were significant for patients' survival outcomes.[30 31] Additionally, other details about treatment regarding surgery, chemotherapy and radiotherapy were unavailable, which should be considered in future research. Finally, the validity of our results still requires external validation using other populations. Meanwhile, more prospective data and other prognostic factors are needed to optimise the accuracy of the nomograms.

CONCLUSION

Nomograms to predict 3-year and 5-year OS and BCSS were constructed and validated based on univariate and multivariate survival analysis. By comparing the

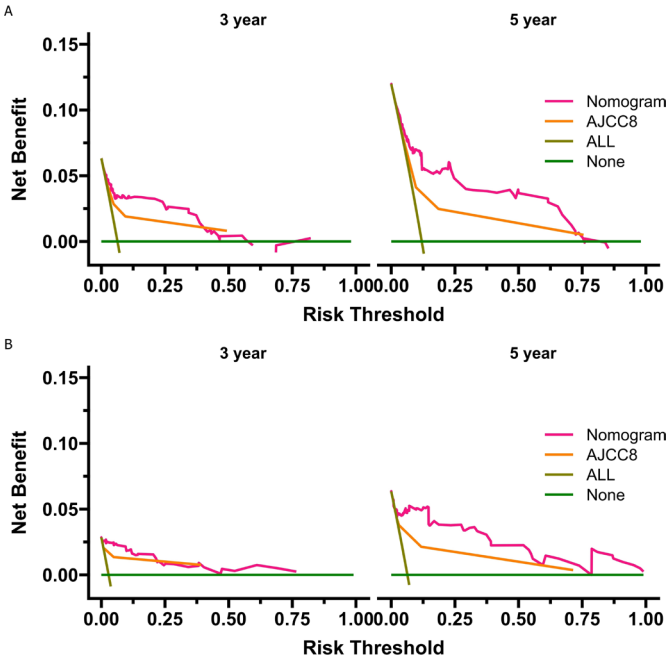

Figure 7 Decision curve analysis of American Joint Committee on Cancer (AJCC) 8 stage and the nomogram. (A) Decision curve analysis 3-year and 5-year overall survival in the validation cohorts. (B) Decision curve analysis 3-year and 5-year breast cancer-specific survival in the validation cohorts.

performances of the nomograms with those of the traditional AJCC staging system, we showed that the nomograms had excellent discrimination and clinical efficacies. These nomograms could identify high-risk patients and thus provide personalised treatment strategies to patients with IMPC.

Contributors YC and PZ designed and conducted the study. YH and ZZ provide suggestions in revision. RT analysed the data. FC drafted the manuscript. J-YS and ZH were responsible for the critical revision. All authors have approved the final version of the manuscript. HZ responsible for the overall content as the guarantor.

Funding This study was supported by the National Natural Science Foundation of China (No. 81872459), the Natural Science Foundation of Guangdong Province (No. 2018A030313666) and the Guangdong Medical Science and Technology Research Fund A2020516.

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not applicable.

Ethics approval All the data used in the present study were taken from the Surveillance, Epidemiology, and End Results database; therefore, local ethics approval was not required.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available upon reasonable request. The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID iD

Yixin Cheng http://orcid.org/0000-0003-4545-5829

REFERENCES

1 Fisher ER, Palekar AS, Redmond C, *et al*. Pathologic findings from the national surgical adjuvant breast project (protocol no. 4). VI. invasive papillary cancer. *Am J Clin Pathol* 1980;73:313–22.
2 Siriaunkgul S, Tavassoli FA. Invasive micropapillary carcinoma of the breast. *Mod Pathol* 1993;6:660–2.
3 Li W, Han Y, Wang C, *et al*. Precise pathologic diagnosis and individualized treatment improve the outcomes of invasive micropapillary carcinoma of the breast: a 12-year prospective clinical study. *Mod Pathol* 2018;31:956–64.
4 Böcker W. WHO classification of breast tumors and tumors of the female genital organs: pathology and genetics. *Verh Dtsch Ges Pathol* 2002;86:116–9.
5 Yang Y-L, Liu B-B, Zhang X, *et al*. Invasive micropapillary carcinoma of the breast: an update. *Arch Pathol Lab Med* 2016;140:799–805.
6 Tang S-L, Yang J-Q, Du Z-G, *et al*. Clinicopathologic study of invasive micropapillary carcinoma of the breast. *Oncotarget* 2017;8:42455–65.
7 Nassar H, Wallis T, Andea A, *et al*. Clinicopathologic analysis of invasive micropapillary differentiation in breast carcinoma. *Mod Pathol* 2001;14:836–41.
8 Kattan MW, Hess KR, Amin MB, *et al*. American joint Committee on cancer acceptance criteria for inclusion of risk models for individualized prognosis in the practice of precision medicine. *CA Cancer J Clin* 2016;66:370–4.
9 Carey LA, Perou CM, Livasy CA, *et al*. Race, breast cancer subtypes, and survival in the Carolina breast cancer study. *JAMA* 2006;295:2492–502.
10 DeSantis CE, Ma J, Goding Sauer A, *et al*. Breast cancer statistics, 2017, racial disparity in mortality by state. *CA Cancer J Clin* 2017;67:439–48.
11 Zhai Z, Zhang F, Zheng Y, *et al*. Effects of marital status on breast cancer survival by age, race, and hormone receptor status: a population-based study. *Cancer Med* 2019;8:4906–17.
12 Wang S, Zhang Y, Yin F, *et al*. Survival outcomes after breast-conserving therapy compared with mastectomy for patients with early-stage invasive micropapillary carcinoma of the breast: a seer population-based study. *Front Oncol* 2021;11:741737.
13 Xiong Z, Deng G, Huang X, *et al*. Score for the survival probability in metastasis breast cancer: a nomogram-based risk assessment model. *Cancer Res Treat* 2018;50:1260–9.
14 Balachandran VP, Gonen M, Smith JJ, *et al*. Nomograms in oncology: more than meets the eye. *Lancet Oncol* 2015;16:S1470-2045(14)71116-7:e173–80.:.
15 Data from: the surveillance, epidemiology, and end results (SEER) database. version 8.3.8.2. n.d. Available: https://seer.cancer.gov/
16 Chen L, Fan Y, Lang R, *et al*. Breast carcinoma with micropapillary features: clinicopathologic study and long-term follow-up of 100 cases. *Int J Surg Pathol* 2008;16:155–63.
17 Assari S, Boyce S, Bazargan M, *et al*. Maternal education at birth and youth breakfast consumption at age 15: blacks' diminished returns. *J (Basel)* 2020;3:313–23.
18 Vingiani A, Maisonneuve P, Dell'orto P, *et al*. The clinical relevance of micropapillary carcinoma of the breast: a case-control study. *Histopathology* 2013;63:217–24.
19 Hao S, Zhao Y-Y, Peng J-J, *et al*. Invasive micropapillary carcinoma of the breast had no difference in prognosis compared with invasive ductal carcinoma: a propensity-matched analysis. *Sci Rep* 2019;9:286.
20 Chen AC, Paulino AC, Schwartz MR, *et al*. Population-Based comparison of prognostic factors in invasive micropapillary and invasive ductal carcinoma of the breast. *Br J Cancer* 2014;111:619–22.
21 Sørlie T, Perou CM, Tibshirani R, *et al*. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A* 2001;98:10869–74.

22 Martínez ME, Unkart JT, Tao L, *et al*. Prognostic significance of marital status in breast cancer survival: a population-based study. *PLoS One* 2017;12:e0175515.
23 Weiss A, Chavez-MacGregor M, Lichtensztajn DY, *et al*. Validation study of the American joint Committee on cancer eighth edition prognostic stage compared with the anatomic stage in breast cancer. *JAMA Oncol* 2018;4:203–9.
24 Ye F-G, Xia C, Ma D, *et al*. Nomogram for predicting preoperative lymph node involvement in patients with invasive micropapillary carcinoma of breast: a seer population-based study. *BMC Cancer* 2018;18:1085.
25 Goutsouliak K, Veeraraghavan J, Sethunath V, *et al*. Towards personalized treatment for early stage HER2-positive breast cancer. *Nat Rev Clin Oncol* 2020;17:233–50.
26 Wu S-G, Zhang W-W, Sun J-Y, *et al*. Postoperative radiotherapy for invasive micropapillary carcinoma of the breast: an analysis of surveillance, epidemiology, and end results database. *Cancer Manag Res* 2017;9:453–9.
27 Meng X, Ma H, Yin H, *et al*. Nomogram predicting the risk of locoregional recurrence after mastectomy for invasive micropapillary carcinoma of the breast. *Clin Breast Cancer* 2021;21:e368–76.
28 Li D, Zhong C, Cheng Y, *et al*. A competing nomogram to predict survival outcomes in invasive micropapillary breast cancer. *J Cancer* 2019;10:6801–12.
29 Denkert C, Seither F, Schneeweiss A, *et al*. Clinical and molecular characteristics of HER2-low-positive breast cancer: pooled analysis of individual patient data from four prospective, neoadjuvant clinical trials. *Lancet Oncol* 2021;22:1151–61.
30 Kurozumi S, Joseph C, Sonbul S, *et al*. A key genomic subtype associated with lymphovascular invasion in invasive breast cancer. *Br J Cancer* 2019;120:1129–36.
31 Buus R, Sestak I, Kronenwett R, *et al*. Molecular drivers of Oncotype DX, prosigna, EndoPredict, and the breast cancer index: a transatac study. *J Clin Oncol* 2021;39:126–35.