Research Paper

Nomogram predicts survival benefit from preoperative radiotherapy for non-metastatic breast cancer: A SEER-based study

Jianjun Liu^{1,*}, Mingxue Su^{2,*}, Shikai Hong^{1,*}, Hong Gao¹, Xucai Zheng¹ and Shengying Wang¹

¹Department of Head, Neck, and Breast Surgery, Anhui Provincial Cancer Hospital, West Branch of Anhui Provincial Hospital, Hefei, China

²Department of Infectious Disease Epidemiology, Lu'an People's Hospital, Lu'an, China

^{*}These authors contributed equally to this work

Correspondence to: Shengying Wang, email: Shengywang@163.com

Keywords: nomogram, breast cancer, preoperative radiotherapy

Received: February 22, 2017 **Accepted:** May 01, 2017 **Published:** May 18, 2017

Copyright: Liu et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License 3.0 (CC BY 3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

ABSTRACT

Background: To estimate survival in non-metastatic breast cancer patients who failed to achieve a pathological complete response (pCR) more effectively, we combined the clinicpathological characteristics after preoperative radiation therapy (pRT) and established a novel nomogram.

Materials and Methods: Using the Surveillance, Epidemiology, and End Results (SEER) database, we identified 2,545 non-metastatic breast cancer patients who underwent pRT between 1998 and 2013. Based on the registries of patients, the primary cohort divided into training set (n = 1,692) and validation set (n = 853). Nomograms were established by training set and validated by validation set.

Results: According to the multivariate analysis of training set, nomogram which combined age at diagnosed, marital status, location, grade, ER status, yp-T status, yp-N status and whether received breast conservation surgery (BCS) was developed. Calibration plots of the nomograms showed that the probability of DSS corresponded to actual observation closely. The C-index was 0.78 in validation set, which was significantly higher than that of yp-TNM staging system (0.75, p = 0.004).

Conclusions: The proposed nomogram resulted in more-reliable DSS prediction for non-metastatic breast cancer patients in general population, it would be helpful in individualized survival prediction and better treatment allocation after pRT.

INTRODUCTION

Breast cancer is one of the most frequently diagnosed cancer in Western country. In 2016, it was estimated that more than 249,000 new breast cancer patients was diagnosed, and it accounted for nearly 30% of all incident cancer in females. In addition, the breast cancer is an aggressive cancer with poor survival, especially in the patients with locally advanced stage or distant metastasis [1]. Neoadjuvant therapy (NAT) may be the only potential curative treatment for locally advanced,

non-distant metastatic breast cancer patients [2, 3]. Furthermore, several studies have demonstrated that NAT can induce tumor-downstaging, thereby promote breast conservation surgery rate for non-metastatic breast cancer patients [4, 5].

Moreover, its can reduce tumor burden, as well as provide a unique opportunity to estimate the outcome of patients by the postoperative pathology [6]. The findings of numerous previous studies have demonstrated that the pCR rate was a strong and acceptable predictor for breast cancer [6–8]. Based on the status of response for NAT, the patients

can be divided into two groups, pCR group and non-pCR group. However, there were only few patients that could achieve pCR, more than 80% patients were diagnosed as tumor residual [9]. In fact, this classification oversimplified the different prognostic categories for the breast cancer patients, especially for the patients with tumor residual.

Currently, the tumor residual burden is evaluated by AJCC stage system, this staging system assumes that tumor cells spread sequentially, which means tumor cells spread firstly from the primary site to lymphatic system, and then translate to distant organs. [10] This strategy solely depends on the final pathologic stage for stratifying patients. However, due to the tumor-downstaging of NAT, this classification may not be applicable for the patients after NAT. In addition, patients' survival was also effected by other individual factors. It is believed that host status and other prognostic factors such as age, race, histology and different treatments after surgery could significantly affect the individual survival in some tumors. [11–13]. Thus, a more accurate survival prediction model incorporating more individual factors was needed.

Recently, Nomogram, a simple predictive tool, have been constructed in several tumors and proved to be useful and effective. [14–18] However, the nomogram applicable for breast cancer after pRT has not been constructed. In this study, we aimed to develop and validate a nomogram based on a multi–institution and multi–population data from SEER database to estimate the survival of patients with non-metastatic breast cancer after pRT.

RESULTS

Clinicopathologic characteristics of patients and follow-up

In primary cohort, a total of 2,545 non-metastatic breast cancer patients who received pRT were included. The clinicopathologic characteristics of breast cancer patients form the training set (n = 1,692) and validation set (n = 853) were listed in Table 1, respectively. Overall, the mean age at diagnosed was 57.3 ± 13.3 years. The majority race of patients was white (84.0%). The majority surgery after radiotherapy was breast conservation surgery (65.9%). 59.7% patients with pathological lymph node-negative. Additionally, 32.0% patients received radiotherapy after surgery (ART). The median follow-up was 95.9 months. 848 (33.3%) patients died before the analysis of the present study. The 3-year, 5-year, 10-year DSS were 88.7%, 84.0% and 75.7%, respectively.

Independent risk factors associated with survival in training set and nomogram development

According to Harrell's guidelines, Clinical pathological variables should been transformed and examined to fit the

Cox PH regression and linear assumption before models construction [19]. The continuous variables were translated into category variables by X-tile software. The best cutoff points for age at diagnosed were 64 and 75 years old (Supplementary Figure 1). In the univariate analysis, most clinicopathologic characteristics, such as Age at diagnosed, marital status, race, grade, tumor location, yp-T stage, yp-N stage, ER status, PR status, BCS and ART, were associated with survival. All the significant factors were included into multivariate analysis.

The result of multivariate analysis was listed in Table 2. Based on the result of multivariate analysis, nomogram predicting 1-year, 3-year and 5-year DSS was constructed (Figure 1).

Nomogram validation

The nomogram was initially validated by bootstrap validation, and then Cross-validated by validation set. As the bootstrap validation result shown, the nomogram model demonstrated more-accuracy for predicting DSS, with an unadjusted C-index of 0.80 and a bootstrap-adjusted C-index of 0.79, which was higher than that of yp-TNM staging system (0.75). The result was similar to the cross-validation. The C-index of the DSS-model was 0.78 (95%CIs, 0.75–0.80), which was significant higher than yp-TNM staging system (0.75, 95%CIs, 0.72–0.78, p = 0.004).

The calibration curves for 1-year, 3-year and 5-year DSS were performed in validation set, and the plots showed that DSS corresponded closely to the actual survival estimated by the Kaplan–Meier method in the validation set (Figure 2). In addition, by the AUC (area of ROC curve) analysis, we also compared the DSS predicting ability of the two models in each time points (Figure 3). As shown in the Figure 3, the proposed nomogram shows more accurate survival predictive ability than the yp-TNM staging system in 3-year and 5-year DSS predicting.

DISCUSSION

In the present study, a total of 2,454 non-metastatic breast cancer patients undergoing pRT from SEER database were analyzed. Based on the clinicopathologic characteristics, we first developed and validated a nomogram to estimate the probability DSS, which exhibited more accuration survival prediction than the yp-TNM staging system.

At present, the models based on postoperative pathology characteristics to predict survival for breast cancer underwent preoperative therapy were still controversy. In 2008, Guarneri et al identified that lymph node status and Ki-67 were the only independent risk factors for breast cancer patients underwent neoadjuvant chemotherapy (NAC) [20]. Ki-67 expression more than 15%, and lymph node metastasis were associated with

Characteristic NO. % NO. % Age (vero) SS 0 + 13.3 55.9 + 13.2 Range 25 to 93 25 to 96 Sex Math 5 0.3 3 0.4 Famale 1687 99.7 850 99.6 Marikal 222 27.2 Marial 22 22.2 27.2 Range 22.2 27.2 No 442 26.1 22.2 27.2 Range 27.2 27		Training set $(n = 1.692)$		Validation set $(n = 853)$	
Age (vars) 55.9 + 13.3 55.9 + 13.2 Mean 55.0 + 13.3 25.0 96 Sex 25.0 93 25.0 96 Maile 5 0.3 3 0.4 Fennle 1687 97.7 850 99.6 Marial T 78 73.9 621 72.8 No 442 26.1 232.2 27.2 Race Tomo focular 659 77.3 Black 157 9.3 119 14.0 APP(7) 8.6 3.3 7.5 8.8 Tomor focular 70 20 5.4 71 8.3 40.6 10.9 10.7 10.0 10.7 10.0 10.7 10.0 10.7 10.0 10.7 10.0 10.7 10.0 10.7 10.0 10.7 10.0 10.7 10.0 10.7 10.0 10.7 10.0 10.7 10.0 10.7 10.0 10.7 10.0 10.7 10.0 10.7 10.0	Characteristic	NO.	%	NO.	%
Sk0 = 13.3 55 9 = 13.2 Range 25 to 93 25 to 96 Sex	Age (years)				
Range25 to 9325 to 90Set50.330.4Fernule168799.785099.6Marital15073.96.2172.8Yes125073.96.2172.8No44226.122.272.7Ree14796.311914.0Aplita1579.311914.0Aplita1579.311914.0Aplita1579.311914.0Aplita1579.311914.0Aplita1579.311914.0Aplita1579.311914.0Aplita1579.311914.0Aplita1579.311914.0Aplita1206.6576.7Operivator1026.6576.7Overlapping lesion1226.6576.7Overlapping lesion1226.6576.7Overlapping lesion2817.110612.4Moderately2817.110612.4Valita10063.841840.0110063.311413.0110063.311413.011046114416.911046311413.01104631313.111046313.013.11 </td <td>Mean</td> <td colspan="2">58.0 ± 13.3</td> <td colspan="2">55.9 ± 13.2</td>	Mean	58.0 ± 13.3		55.9 ± 13.2	
Sx Sx S <ths< th=""> S S S</ths<>	Range	25 to 93		25 to 96	
Male 5 0.3 3 0.4 Female 1687 99.7 850 99.6 Marial	Sex				
Fende168799.785099.6Marital	Male	5	0.3	3	0.4
Marial Yes 1250 7.3 6.21 7.2.8 No 442 2.61 2.32 7.2.8 Whe 1479 7.4 6.69 7.7.3 Black 157 9.3 119 14.0 API/A1 6.59 7.3 8.0 7.0 Black 157 9.3 119 14.0 API/A1 6.5 3.3 7.5 8.8 CenterNipple 92 5.4 7.1 8.3 Upper-outer 7.11 4.56 5.7 6.7 Lower-outer 138 8.2 6.0 7.0 6.7 Lower-outer 138 8.2 6.0 7.0 6.7 Overlapping lesion 7.9 7.2 6.7 6.7 6.7 Moderatory 7.68 5.8 4.18 49.0 Undifferentiated 2.8 1.7 1.4 1.6 Syster 6.6 5.8 4.18 <	Female	1687	99.7	850	99.6
Yes 1250 73.9 621 72.8 No 420 26.1 232 22.7 No 1479 87.4 6.59 77.3 Black 157 9.3 119 14.0 API/A1 56 3.3 75 8.8 Tumor location - - - 8.8 Qupor-outer 77.1 45.6 34.6 40.6 Upper-imer 122 5.4 71 8.3 Upper-imer 138 8.2 60 70 Lower-outer 138 8.2 60 70 Corderping lesion 379 2.2.4 282 26.7 Grade	Marital				
No 442 26.1 232 27.2 Race	Yes	1250	73.9	621	72.8
Race Number 1479 87.4 659 77.3 Black 157 9.3 119 14.0 API/AI 56 3.3 75 8.8 Tumor location E E E E Upper-inter 77.1 4.5.6 34.6 40.6 Upper-inter 200 11.8 9.1 10.7 Lower-outer 138 8.2 60 7.0 Lower-outer 138 8.2 60 7.0 Coverlapping lesion 379 2.2.4 2.28 2.67 Grade 12 66 57 6.7 6.7 Well 200 17.1 106 1.4 4.0 Moderately 768 45.4 315 36.9 Party 12 12 60 3.48 40.8 2 462 27.3 250 29.3 3 101 6.1 141 16.9 <t< td=""><td>No</td><td>442</td><td>26.1</td><td>232</td><td>27.2</td></t<>	No	442	26.1	232	27.2
White 1479 87.4 659 77.3 Black 157 9.3 119 14.0 API/AI 56 3.3 75 8.8 Tumor location 8.3 Center/Nipple 92 5.4 71 8.3 Upper-outer 771 45.6 346 40.6 Upper-outer 120 6.6 57 6.7 Lower-outer 138 8.2 60 7.0 Grader 666 35.8 418 49.0 Undifferentiated 28 1.7 14 16 Poorly 606 35.8 418 49.0 Undifferentiated 28 1.7 14 16 Porty 606 35.8 418 49.0 Undifferentiated 28 1.7 14 16 Sym <tage< td=""> 102 6.3 1111 13.0 4 104 6.1</tage<>	Race				
Black 157 9.3 119 14.0 API/AI 56 3.3 75 8.8 Tumor location	White	1479	87.4	659	77.3
API/AI 56 3.3 75 8.8 Tumo Icoation	Black	157	9.3	119	14.0
Tumor locationInInInCentr/Nipple925.47.18.3Upper-outer77145.634.640.6Upper-outer20011.89.110.7Lower-outer1388.26.07.0Lower-outer1126.65.76.7Overlapping lesion37.92.2.42.2826.7Grade	API/AI	56	3.3	75	8.8
Center/Nipple 92 5.4 71 8.3 Upper-outer 771 45.6 346 40.6 Upper-outer 771 45.6 346 40.6 Upper-outer 138 8.2 60 7.0 Lower-outer 138 8.2 60 7.0 Overlapping lesion 379 2.24 228 2.67 Grade 7.1 10.6 12.4 Well 290 17.1 10.6 12.4 Moderately 788 45.4 315 36.9 Poorly 606 35.8 418 49.0 Undifferentiated 28 1.7 14 1.6 Yp-T stage 1 1.6 1.0 1.0 1 1020 60.3 348 40.8 2 462 27.3 250 29.3 3 106 64.5 30.8 12.7 59 56 53	Tumor location				
$L_{perimer}$ 77145.634640.6Upper-inner20011.89110.7Lower-outer1388.2607.0Lower-inner1126.6576.7Overlapping lesion3792.242.2826.7Overlapping lesion3792.242.282.6.7Overlapping lesion3792.242.282.6.7Overlapping lesion3792.242.282.6.7Overlapping lesion3792.242.282.6.7Well29017.110.61.2.4Moderately7684.5.43.153.6.9Poorly0603.5.84.1849.0Undifferentiated2.81.71.41.6Yp-Tatage1102060.33.1113.041046.11.441.01.041046.11.441.01.041046.11.441.01.041046.11.441.01.041.046.11.441.01.041.046.11.441.01.041.046.11.441.01.041.046.11.441.01.051.13.402.0.12.633.0.8213.402.0.12.633.0.821.679.91.081.2.7 <t< td=""><td>Center/Nipple</td><td>92</td><td>5.4</td><td>71</td><td>8.3</td></t<>	Center/Nipple	92	5.4	71	8.3
Upper-inner 200 11.8 91 10.7 Lower-outer 138 8.2 60 7.0 Lower-outer 138 8.2 60 7.0 Lower-outer 137 22.4 228 26.7 Grade 768 45.4 315 36.9 Poorly 606 35.8 418 49.0 Undifferentiated 28 1.7 14 1.6 Yp-T stage 1020 60.3 348 40.8 29.3 3 1.06 6.3 111 13.0 4 10.9 1.0 4.0 10.9 1.0 4.0 10.9 1.0 </td <td>Upper-outer</td> <td>771</td> <td>45.6</td> <td>346</td> <td>40.6</td>	Upper-outer	771	45.6	346	40.6
Lower-outer 138 8.2 60 7.0 Lower-inner 112 6.6 57 6.7 Overlapping lesion 379 22.4 228 26.7 Grade	Upper-inner	200	11.8	91	10.7
Lover-inner 112 6.6 57 6.7 Overlapping lesion 379 22.4 228 26.7 Grade	Lower-outer	138	8.2	60	7.0
Overlapping lesion3792.42.2826.7Grade $=$ <	Lower-inner	112	6.6	57	6.7
Grade Lin Lin <thlin< th=""> <thlin< td="" th<=""><td>Overlapping lesion</td><td>379</td><td>22.4</td><td>228</td><td>26.7</td></thlin<></thlin<>	Overlapping lesion	379	22.4	228	26.7
Well29017.110612.4Moderately76845.431536.9Poorly60635.841849.0Undifferentiated281.7141.6Yp-T stage1102060.334840.8227.325029.331066.311131066.311113.0416.9yp-N stage.11046.114416.9yp-N stage.134020.126.330.821679.910812.73945.65.36.2ER status10565.345.656.1P status10565.345.656.1PR status10565.345.656.1P status10565.345.6 <td>Grade</td> <td></td> <td></td> <td></td> <td></td>	Grade				
Moderately 768 454 15 16.9 Poorly 606 35.8 418 49.0 Undifferentiated 28 1.7 14 1.6 Yp-T stage 1 1020 60.3 348 40.8 2 462 27.3 250 29.3 3 106 6.3 111 13.0 4 104 6.1 144 16.9 yp-N stage. 0 104 6.1 144 16.9 yp-N stage. 0 107 9.9 108 12.7 3 94 5.6 53 6.2 ER status 12.7 3.9 9.4 5.6 53 6.2 Positive 167 9.9 108 12.7 3 9.4 5.6 53 6.2 ER status 1105 5.6 53 6.2 5.1 13.1 15.1 Ygative 587 34.7 389	Well	290	171	106	12.4
Poorty For FA FA FA PA Poorty 606 3.8 4.18 49.0 Undifferentiated 2.8 1.7 1.4 1.6 Yp-T stage 3 40.8 40.8 2 462 27.3 250 29.3 3 106 6.3 111 13.0 4 104 6.1 144 16.9 yp-N stage. 340 20.1 263 30.8 2 167 9.9 108 12.7 3 2 30.8 2.7 3 94 5.6 53 6.2 2.8 3.9 3.9 3.9 Positive 167 9.9 108 12.7 3.9 6.6 5.1 2.9 3.9 5.6 5.1 7.7 3.9 6.5 1.2 7.7 3.9 5.6 5.1 7.9 1.0 5.6 5.1 7.9 1.0 <td>Moderately</td> <td>768</td> <td>45.4</td> <td>315</td> <td>36.9</td>	Moderately	768	45.4	315	36.9
Lock Lock Lock Lock Lock Lock Lock Lock Undifferentiated 28 1.7 14 1.6 Yp-T stage 1 1020 60.3 348 40.8 2 462 27.3 250 29.3 3 106 6.3 111 13.0 4 104 6.1 144 16.9 yp-N stage. 0 1091 64.5 429 50.3 1 340 20.1 263 30.8 2.7 3 340 20.1 263 30.8 2.7 3 340 20.1 263 30.8 12.7 3 340 20.1 263 30.8 12.7 3 94 5.6 53 6.2 12.7 3 167 9.9 108 12.7 3 1128 75.9 564 56.1 Prestive 1105	Poorly	606	35.8	418	49.0
Typ-T stage In In <thin< th=""> In In</thin<>	Undifferentiated	28	17	14	1.6
1 1020 60.3 348 40.8 2 462 27.3 250 29.3 3 106 6.3 111 13.0 4 104 6.1 144 169 yp-N stage. 0 1091 64.5 429 50.3 1 340 20.1 263 30.8 2 167 9.9 108 12.7 3 94 5.6 53 6.2 Restatus Negative 408 24.1 289 33.9 Positive 1284 75.9 564 56.1 PR status 34.7 389 45.6 Positive 1105 65.3 451 52.9 Unknown 1105 65.3 13 1.5 <td>Yn-T stage</td> <td></td> <td>**/</td> <td></td> <td>1.0</td>	Yn-T stage		**/		1.0
1 100 0.00 0.00 0.00 0.00 2 462 27.3 250 29.3 3 106 6.3 111 13.0 4 104 6.1 144 16.9 yp-N stage. 50.5 50.3 0 1091 64.5 429 50.3 50.3 50.5 50.3 50.5 50.3 50.5 50.3 50.5 50.3 50.5 50.3 50.5 50.3 50.5 50.3 50.5 50.3 50.5	1	1020	60.3	348	40.8
100 100 100 100 100 3 104 6.1 114 16.9 yp-N stage. 0 1091 64.5 429 50.3 1 340 20.1 263 30.8 2 167 9.9 108 12.7 3 94 5.6 53 6.2 ER status Negative 408 24.1 289 33.9 Positive 1284 75.9 564 56.1 PR status Negative 567 34.7 389 45.6 Positive 1105 65.3 13 1.5 yp-AJCC Stage 110 495 29.3 290 34.0 II 827 48.9 261 30.6 30.6 II 495 29.3 290 34.0 34.0 III 495 29.3 290 34.0 34.0 II 495 29.3 396	2	462	27.3	250	29.3
A 100 6.5 111 16.6 4 104 6.1 144 16.9 yp-N stage. 0 1091 64.5 429 50.3 1 340 20.1 263 30.8 2 167 9.9 108 12.7 3 94 5.6 53 6.2 ER status Negative 408 24.1 289 33.9 Positive 1284 75.9 564 56.1 PR status Negative 57 34.7 389 45.6 Positive 1105 65.3 13 1.5 yp-AJCC Stage I 827 48.9 261 30.6 II 495 29.3 290 34.0 III	3	106	63	111	13.0
yp-N stage. initial initial initial 0 1091 64.5 429 50.3 1 340 20.1 263 30.8 2 167 9.9 108 12.7 3 94 5.6 53 6.2 ER status	4	104	6.1	144	16.9
Instruct Instruct	vn-N stage	101	0.1	111	10.9
0 101 013 013 125 303 1 340 20.1 263 30.8 2 167 9.9 108 12.7 3 94 5.6 53 6.2 ER status	0	1091	64 5	429	50.3
1 100 101 100 100 2 167 9.9 108 12.7 3 94 5.6 53 62 ER status	1	340	20.1	263	30.8
2 101 102 103 105 65.3 13 1.5 103 105 65.3 13 1.5 105 105 65.3 13 1.5 105 105 65.3 13 1.5 105 105 65.3 13 1.5 105	2	167	9.9	108	12.7
ER status 6.0 6.0 6.1 Negative 408 24.1 289 33.9 Positive 1284 75.9 564 56.1 PR status 57.0 564 56.1 PR status 57.0 564 56.1 PR status 389 45.6 Positive 1105 65.3 451 52.9 Unknown 1105 65.3 13 1.5 yp-AJCC Stage 37.0 29.0 34.0 II 827 48.9 261 30.6 30.6 III 370 29.3 290 34.0 III 370 21.9 302 35.4 BCS 37.6 33.6 No 473 28.0 396 46.4 ART 37.6 Yes 518 30.6 297 34.8 No 1174 69.4 556 65.2	3	94	5.6	53	62
Negative 408 24.1 289 33.9 Positive 1284 75.9 564 56.1 PR status	FR status		5.0	55	0.2
Positive 100 241 269 563 Positive 1284 75.9 564 56.1 PR status	Negative	408	24.1	289	33.9
PR status res best best Negative 587 34.7 389 45.6 Positive 1105 65.3 451 52.9 Unknown 1105 65.3 13 1.5 yp-AJCC Stage I 827 48.9 261 30.6 II 827 48.9 261 30.6 III 495 29.3 290 34.0 III 370 21.9 302 35.4 BCS Ves 1219 72.0 457 53.6 No 473 28.0 396 46.4 ART Ves 518 30.6 297 34.8 No 1174 69.4 556 65.2	Positive	1284	75.9	564	56.1
Negative 587 34.7 389 45.6 Positive 1105 65.3 451 52.9 Unknown 1105 65.3 13 1.5 yp-AJCC Stage I 827 48.9 261 30.6 II 495 29.3 290 34.0 III 370 21.9 302 35.4 BCS Ves 1219 72.0 457 53.6 No 473 28.0 396 46.4 ART Yes 518 30.6 297 34.8 No 1174 69.4 556 65.2	PR status				
Positive 1105 65.3 451 52.9 Unknown 1105 65.3 13 1.5 yp-AJCC Stage I 827 48.9 261 30.6 II 495 29.3 290 34.0 III 370 21.9 302 35.4 BCS Yes 1219 72.0 457 53.6 No 473 28.0 396 46.4 ART Yes 518 30.6 297 34.8 No 1174 69.4 556 65.2	Negative	587	34 7	389	45.6
Hourse Hou Hou <t< td=""><td>Positive</td><td>1105</td><td>65.3</td><td>451</td><td>52.9</td></t<>	Positive	1105	65.3	451	52.9
yp-AJCC Stage No 827 48.9 261 30.6 II 495 29.3 290 34.0 III 370 21.9 302 35.4 BCS 1219 72.0 457 53.6 No 473 28.0 396 46.4 ART Yes 518 30.6 297 34.8 No 1174 69.4 556 65.2	Unknown	1105	65.3	13	15
I 827 48.9 261 30.6 II 495 29.3 290 34.0 III 370 21.9 302 35.4 BCS 72.0 457 53.6 No 473 28.0 396 46.4 ART 72.8 518 30.6 297 34.8 No 1174 69.4 556 65.2	vn-AICC Stage	1100	00.0	10	1.0
II 495 29.3 290 34.0 III 370 21.9 302 35.4 BCS 788 1219 72.0 457 53.6 No 473 28.0 396 46.4 ART 728 518 30.6 297 34.8 No 1174 69.4 556 65.2	I	827	48.9	261	30.6
III 370 21.9 302 35.4 III 370 21.9 302 35.4 BCS	I	495	29 3	290	34.0
BCS Yes 1219 72.0 457 53.6 No 473 28.0 396 46.4 ART Yes 518 30.6 297 34.8 No 1174 69.4 556 65.2	 III	370	21.9	302	35.4
Yes 1219 72.0 457 53.6 No 473 28.0 396 46.4 ART	BCS	570	/	202	55.1
No 473 28.0 396 46.4 ART	Yes	1219	72.0	457	53.6
ART Yes 518 30.6 297 34.8 No 1174 69.4 556 65.2	No	473	28.0	396	46.4
Yes 518 30.6 297 34.8 No 1174 69.4 556 65.2	ART		20.0	570	10.7
No 1174 694 556 652	Yes	518	30.6	297	34.8
	No	1174	69.4	556	65.2

Table 1: Characteristic of training set and validation set

Abbreviation: API, Asian or Pacific Islander; AI, Asian or Pacific Islander; ER status: Estrogen Receptor status; PR status: Progesterone Receptor status; BCS, Breast Conservation Surgery; ART, Adjuvant radiotherapy; AJCC, American Joint Committee on Cancer.

	HR	95% CI	р
Age (years)			0.017
< 64	ref		
65~75	1.22	0.93 to 1.61	
> 76	1.62	1.14 to 2.30	
Race			
White			
Black			
API/AI			
Marital			< 0.001
Yes	ref		
No	0.65	0.52 to 0.83	
Location			0.018
Center/Nipple	ref		
Upper- inner	1.15	0.65 to 2.01	
Upper- outer	1.28	0.81 to 2.02	
Overlapping lesion	1.37	0.85 to 2.21	
Lower- inner	1.43	0.81 to 2.54	
Lower- outer	2.38	1.37 to 4.14	
Grade			0.007
Well	ref		
Moderately	1.44	0.92 to 2.27	
Poorly	1.90	1.20 to 3.020	
Undifferentiated	2.72	1.35 to 5.48	
vp-T stage			< 0.001
1	ref		
2	1.75	1.32 to 2.31	
3	2.23	1.45 to 3.44	
4	3.39	2.28 to 5.04	
vp-N stage			< 0.001
0	ref		
1	1.93	1.43 to 2.60	
2	3.62	2.59 to 5.05	
3	4.75	3.25 to 6.94	
ER status			< 0.001
Positive	ref		
Negative	0.54	0.43 to 0.69	
BCS			0.011
Yes	ref		
No	0.68	0.50 to 0.91	

Table 2: Multivariate analysis of the training set

Abbreviation: HR: Hazard Ratio; API, Asian or Pacific Islander; AI, Asian or Pacific Islander; ER status: Estrogen Receptor status; BCS, Breast Conservation Surgery.

worse DSS (HR = 3.75, P < 0.001, HR = 2.31, P = 0.037, respectively). Based on those findings, the authors proposed a new classification including Ki-67 and lymph node status for breast cancer patients underwent NAC. Those findings were interesting, however, the sample size of that study was small. The selection bias may exist. Furthermore, several other important clinicopathologic characteristics had been ignored. For example, the tumor location, patients' age at diagnosed and the postoperative

treatment, which may also affect patients' survival. It is believed that a more refined prognostic classification which included more risk factors for breast cancer patients was needed.

To date, several nomograms had been constructed, and shown more accurate survival prediction in breast cancer patients underwent NAC. In 2005, based on 1,070 breast cancer patients with pCR evaluation forming various clinical trials, Rouzier and colleagues developed



Figure 1: Nomogram predicting 1–year, 3–year and 5–year DSS for non-metastasis breast cancer patients after pRT. The nomogram was used by accumulating the points identified on the points scale for each variable. Based the sum of these points projected on the bottom scales, the nomogram can provide the probability of 1–year, 3–year and 5–year DSS for an individual patient. Abbreviation: ER, estrogen receptor; DSS, disease specific survival; BCS, breast conservation surgery; pRT, neoadjuvant radiotherapy; API, Asian or Pacific Islander.



Figure 2: The calibration curve for predicting patients' DSS at 1–year (\mathbf{A}), 3–year (\mathbf{B}) and 5–year (\mathbf{C}) in the validation set. The X–aixs represented the nomogram–predicted survival, and the actual survival was plotted on the Y–axis. The dotted line represented the ideal correlationship between nomogram-predicted and actual survival. Abbreviation: DSS, disease specific survival.

and validated a nomogram, which included clinical stage, ER status, grade and the number of preoperative chemotherapy cycles to estimate distant metastasisfree survival [21]. This nomogram was validated by two independent institution data, and the C-index were both higher than 0.77. This nomogram was important, it was the first nomogram to predict breast cancer patients undergoing preoperative therapy. However, according to Harrell's guidelines, this nomogram was based on the pCR patients, it was not suitable for the patients with tumor residual. In 2011, Keam et al developed a nomogram, which included age at diagnosed, initial clinical stage, yp-TNM stage, ER status and Ki-67. Its C-index was 0.78, and could be used to predict the probability of 2-year relapse-free survival. However, the patients of this study were all received NAC, it might not be suitable for patients undergoing pRT.

In this study, we first analyzed a multi–institution and multi–population database, and identified the risk factors for breast cancer underwent pRT. Besides lymph node status, we identified several other clinicopathologic characteristics associated with patients' survival. Combining all the risk factors, we developed and validated DSS-model to predict breast cancer patients' survival, and the nomogram exhibited more-accuracy survival prediction than yp-TNM staging system. Indeed, it would be helpful to design an individual postoperative treatment and the schedule of follow-up for breast cancer patient.

Although the nomogram demonstrated a moreaccuracy survival prediction, several limitations should not be ignored. Firstly, there may be a selection bias in the primary cohort, since only the patients who had complete information were included in present study. Secondly, as those nomograms were based on SEER database, all the analysis were limited to the prognosis factors in the database. Several predictors such as, Ki-67 index, tumor genetic differences, whether received neoadjuvant chemotherapy, hormone therapy and postoperative chemotherapies had not been analyzed. Lastly, patients included in the study may also had received other therapy, such as NAC and/or preoperative hormone therapy, which may limit the statistics power of this study.

In conclusion, our study first developed and validated a prognostic nomogram based on a multiinstitution and multi-population database predicting DSS for breast cancer patients undergoing pRT. Compared to the yp-TNM staging system, the proposed nomogram represented better prognostic discrimination and predictive accuracy for DSS. It should be helpful to calculate individualized survival prediction and provide better treatment allocation for non-metastasis breast cancer after pRT.

MATERIALS AND METHODS

Patients

The primary cohort was derived from the SEER program. The SEER program is a national collaboration program by the National Cancer Institute of Unit State. Approximately 3 million cases from a variety of geographic regions had been collected and published, and it covers nearly 26% American population's cancer incidence and survival data.

A retrospective review of all breast cancer patients after pRT with tumor residual from SEER database between 1998 and 2013 was performed. A total of 7,310 patients from 18 SEER registries were initially screened. Since the patients with distant metastasis have an obviously worse survival than non-metastasis patients, those patients were not analyzed in current study. In addition, patients were excluded if they had incomplete information(s) on tumor size (yp-T stage), lymph node stage (yp-N stage) or status of distant organs metastasis (M stage). The remained (n = 2,545) were defined as SEER primary cohort. In the SEER





primary cohort, patients from 7 randomly selected registries (Louisiana, New Jersey, New Mexico, Rural Georgia, San Francisco-Oakland SMSA, San Jose-Monterey, Seattle, Utah) were assigned as training data set, and from other registries were regarded as SEER validation set.

Outcomes

The primary endpoint was DSS, which was defined as the time form surgery to cancer–related death or the last follow–up. The follow–up duration was measured as the time from the date of surgery to the last follow–up. The survival status was recorded according to the latest follow–up.

Covariates

The data of patients' clinicpathological characteristics such as age at diagnosed, sex, race, marital, surgery, tumor location, tumor size (continuous variable), histology, grade, yp-T stage (category variable), yp-N stage, estrogen receptor (ER) status, progesterone receptor (PR) status, M stage, number of positive metastasis lymph nodes and number of examined lymph nodes were collected. The pathological tumor stage, yp-T stage and yp-N stage were restaged according to the 7th edition of AJCC staging system.

Risk factors selected and the construction of the nomogram

The method was performed as our previous studies described [22, 23]. Briefly, the nomogram assumes that a linear correlation between risk factors and patients survival [19]. The linear relationships between continuous variables and survival were evaluated by restricted cubic splines. Before modeling, continuous variables were transformed into categorical variables to fit the linear assumption. The best cut-off points of continuous variables were identified by X-tile [24]. Categorical variables were grouped based on clinical findings before modeling. By the forward stepwise in the Cox proportional hazards (PH) regression model, all the independent risk factors were identified. DSS estimation and survival curve were performed by Kaplan–Meier method, and validated by the log–rank test.

The model of nomogram was established based on the data of training set. According to the results of Cox PH regression, nomogram combining all the independent prognostic factors was constructed for predicting 1–year, 3–year and 5–year DSS by using the package of *rms* in R software version 3.2.4.

Validation of the nomogram

The nomogram was validated by measuring both discrimination and calibration in validation dataset. Firstly,

Discrimination between the proposed nomogram and the 7th edition of AJCC staging system were performed by the roccp.cens package in R. Bootstrapping is an internal validation whereby the model is iteratively applied to randomly selected sample sets of the primary cohort. However, external validation is performed in many disparate cohorts, which is the gold standard validation. Therefore, in this part, the nomogram was internal validated by bootstrap in the training set and external vxalidated in the validation set. The discrimination of nomogram was evaluated by Harrell's C-index, which could estimate the probability between the observed and predicted DSS [19]. The higher the C-index, the more precise the survival prediction was. Following, calibrations were firstly carried out by grouping all the patients, and then the mean of the groups were compared with observed Kaplan-Meier DSS estimation.

Statistical analysis

All the p value less than 0.05 will be considered as statistically significant. All statistics analysis were performed by the R software version 3.2.4 (http://www.rproject.org), X-title and the software statistical package for social sciences version 19.0 (SPSS, Chicago, IL).

Ethical approval

Current study does not contain any studies with human participants or animals performed by any of the authors.

Authors' contributions

Conception and design: Shengying Wang, Shikai Hong; Development of methodology: Shengying Wang, Shikai Hong, Xucai Zheng; Collection and assembly of data: Jianjun Liu, Mingxue Su; Data analysis and interpretation: Jianjun Liu, Mingxue Su, Hong Gao; Writing, review, and/or revision of the manuscript: All authors; Administrative, technical, or material support: Shengying Wang, Shikai Hong; Study supervision: Shikai Hong, Xucai Zheng; Final approval of manuscript: All authors

ACKNOWLEDGMENTS

The authors thank all medical personnel of the Department of Head - neck and Thoracic Surgery, Anhui Provincial Cancer Hospital, West branch of Anhui Provincial Hospital, Hefei, China, for their technical assistance.

CONFLICTS OF INTEREST

The authors have no potential conflicts of interest.

GRANT SUPPORT

None.

REFERENCES

- 1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2016. CA Cancer J Clin. 2016; 66:7–30.
- Bonadonna G, Valagussa P, Brambilla C, Ferrari L, Moliterni A, Terenziani M, Zambetti M. Primary chemotherapy in operable breast cancer: eight-year experience at the Milan Cancer Institute. J Clin Oncol. 1998; 16:93–100.
- 3. Untch M, Konecny GE, Paepke S, von Minckwitz G. Current and future role of neoadjuvant therapy for breast cancer. Breast. 2014; 23:526–537.
- Killelea BK, Yang VQ, Mougalian S, Horowitz NR, Pusztai L, Chagpar AB, Lannin DR. Neoadjuvant chemotherapy for breast cancer increases the rate of breast conservation: results from the National Cancer Database. J Am Coll Surg. 2015; 220:1063–1069.
- Golshan M, Cirrincione CT, Sikov WM, Berry DA, Jasinski S, Weisberg TF, Somlo G, Hudis C, Winer E, Ollila DW. Impact of neoadjuvant chemotherapy in stage II-III triple negative breast cancer on eligibility for breast-conserving surgery and breast conservation rates: surgical results from CALGB 40603 (Alliance). Ann Surg. 2015; 262:434–439.
- Cortazar P, Zhang L, Untch M, Mehta K, Costantino JP, Wolmark N, Bonnefoi H, Cameron D, Gianni L, Valagussa P, Swain SM, Prowell T, Loibl S, et al. Pathological complete response and long-term clinical benefit in breast cancer: the CTNeoBC pooled analysis. Lancet. 2014; 384:164–172.
- Cortazar P, Geyer CE Jr. Pathological complete response in neoadjuvant treatment of breast cancer. Ann Surg Oncol. 2015; 22:1441–1446.
- Bonnefoi H, Litiere S, Piccart M, MacGrogan G, Fumoleau P, Brain E, Petit T, Rouanet P, Jassem J, Moldovan C, Bodmer A, Zaman K, Cufer T, et al. Pathological complete response after neoadjuvant chemotherapy is an independent predictive factor irrespective of simplified breast cancer intrinsic subtypes: a landmark and two-step approach analyses from the EORTC 10994/BIG 1–00 phase III trial. Ann Oncol. 2014; 25:1128–1136.
- Bonadonna G, Veronesi U, Brambilla C, Ferrari L, Luini A, Greco M, Bartoli C, Coopmans de Yoldi G, Zucali R, Rilke F, Andreola S, Silvestrini R, Di Fronzo G, Valagussa P. Primary chemotherapy to avoid mastectomy in tumors with diameters of three centimeters or more. J Natl Cancer Inst. 1990; 82:1539–1545.
- Edge S, Byrd DR, Compton CC, Fritz AG, Green F, Trotti A, editors. AJCC Cancer Staging Handbook. 7th Edition. New York, NY: Springer; 2010.
- 11. Kalache A, Maguire A, Thompson SG. Age at last full-term pregnancy and risk of breast cancer. Lancet. 1993; 341:33–36.

- O'Brien KM, Cole SR, Tse CK, Perou CM, Carey LA, Foulkes WD, Dressler LG, Geradts J, Millikan RC. Intrinsic breast tumor subtypes, race, and long-term survival in the Carolina Breast Cancer Study. Clin Cancer Res. 2010; 16:6100–6110.
- Wohlfahrt J, Mouridsen H, Andersen PK, Melbye M. Reproductive risk factors for breast cancer by receptor status, histology, laterality and location. Int J Cancer. 1999; 81:49–55.
- Iasonos A, Schrag D, Raj GV, Panageas KS. How to build and interpret a nomogram for cancer prognosis. J Clin Oncol. 2008; 26:1364–1370.
- Balachandran VP, Gonen M, Smith JJ, DeMatteo RP. Nomograms in oncology: more than meets the eye. The Lancet Oncology. 2015; 16:e173-e180.
- 16. Liang W, Zhang L, Jiang G, Wang Q, Liu L, Liu D, Wang Z, Zhu Z, Deng Q, Xiong X, Shao W, Shi X, He J. Development and validation of a nomogram for predicting survival in patients with resected non-small-cell lung cancer. J Clin Oncol. 2015; 33:861–869.
- Wang Y, Li J, Xia Y, Gong R, Wang K, Yan Z, Wan X, Liu G, Wu D, Shi L, Lau W, Wu M, Shen F. Prognostic nomogram for intrahepatic cholangiocarcinoma after partial hepatectomy. J Clin Oncol. 2013; 31:1188–1195.
- Kattan MW, Karpeh MS, Mazumdar M, Brennan MF. Postoperative nomogram for disease-specific survival after an R0 resection for gastric carcinoma. J Clin Oncol. 2003; 21:3647–3650.
- Harrell FE Jr, Califf RM, Pryor DB, Lee KL, Rosati RA. Evaluating the yield of medical tests. Jama. 1982; 247:2543–2546.
- 20. Guarneri V, Piacentini F, Ficarra G, Frassoldati A, D'Amico R, Giovannelli S, Maiorana A, Jovic G, Conte P. A prognostic model based on nodal status and Ki-67 predicts the risk of recurrence and death in breast cancer patients with residual disease after preoperative chemotherapy. Ann Oncol. 2009; 20:1193–1198.
- Rouzier R, Pusztai L, Delaloge S, Gonzalez-Angulo AM, Andre F, Hess KR, Buzdar AU, Garbay JR, Spielmann M, Mathieu MC, Symmans WF, Wagner P, Atallah D, et al. Nomograms to predict pathologic complete response and metastasis-free survival after preoperative chemotherapy for breast cancer. J Clin Oncol. 2005; 23:8331–\$3339.
- 22. Liu J, Geng Q, Chen S, Liu X, Kong P, Zhou Z, Zhan Y, Xu D. Nomogram based on systemic inflammatory response markers predicting the survival of patients with resectable gastric cancer after D2 gastrectomy. Oncotarget. 2016; 7:37556–37565. doi: 10.18632/oncotarget.8788.
- 23. Liu J, Geng Q, Liu Z, Chen S, Guo J, Kong P, Chen Y, Li W, Zhou Z, Sun X, Zhan Y, Xu D. Development and external validation of a prognostic nomogram for gastric cancer using the national cancer registry. Oncotarget. 2016; 7:35853–35864. doi: 10.18632/oncotarget.8221.
- Camp RL, Dolled-Filhart M, Rimm DL. X-tile: a new bioinformatics tool for biomarker assessment and outcomebased cut-point optimization. Clin Cancer Res. 2004; 10:7252–7259.