



doi:10.7659/j.issn.1005-6947.2020.11.002  
http://dx.doi.org/10.7659/j.issn.1005-6947.2020.11.002  
Chinese Journal of General Surgery, 2020, 29(11):1293-1302.

· 专题研究 ·

## 基于临床病理大数据的早发乳腺癌腋窝淋巴结转移风险预测模型构建

宋文静<sup>1</sup>, 贺鑫<sup>1</sup>, 龚鹏举<sup>1</sup>, 杨燕<sup>1</sup>, 魏蕾<sup>2</sup>, 张京伟<sup>1</sup>

(1. 武汉大学中南医院 甲乳外科, 湖北 武汉 430071; 2. 武汉大学基础医学院 病理与病理生理学教研室, 湖北 武汉 430071)

### 摘要

**背景与目的:** 近年来, 乳腺癌的发病人群趋于年轻化, 并且更容易发生腋窝淋巴结 (ALN) 转移。本研究通过临床病理大数据平台分析年轻乳腺癌患者 ALN 转移的影响因素, 并建立风险预测模型, 为年轻乳腺癌患者的诊断和治疗提供参考依据。

**方法:** 收集 SEER 数据库中 2010—2015 年间被诊断为乳腺癌并且接受了 ALN 手术的年轻患者的临床病理资料, 采用单因素和多因素回归分析筛选 ALN 转移的影响因素, 并以列线图的方式可视化。通过 AUC/C 指数量化列线图区分不同 ALN 状态患者的能力, 采用 bootstrap 方法 (1 000 次重复, 随机数种子设置为 12) 进行列线图预测性能内部验证。另外, 收集 2015—2017 年在武汉大学中南医院初诊为乳腺癌的年轻患者资料, 对模型行外部验证。

**结果:** 共纳入 SEER 数据库中 23 778 例年轻乳腺癌患者, 其中 39.6% 患者存在 ALN 转移。单因素 Logistic 回归分析显示, 年龄、种族、肿瘤部位、病理学分级、肿瘤大小、胸壁或皮肤是否受侵以及 ER、PR、HER-2 状态与 ALN 转移有关 (均  $P < 0.001$ ) ; 多因素 Logistic 回归分析显示: 年龄、种族、婚姻状态、边侧、肿瘤部位、分级、肿瘤大小、胸壁或皮肤是否受侵以及 ER 与 PR 状态是 ALN 转移独立影响因素 (均  $P < 0.05$ ) , 据此建立风险预测模型。内部验证的校准曲线显示, 利用该模型计算的预测值与真实值之间存在良好的一致性 (AUC/C 指数 = 0.716) 。共纳入 391 例年轻乳腺癌患者作为外部验证数据集, 其中 49.9% 患者初次手术发现有 ALN 转移。外部验证提示模型预测能力较好 (AUC/C 指数 = 0.798) 。

**结论:** 基于 SEER 数据库建立的年轻乳腺癌患者 ALN 转移的风险预测模型具有较好的预测能力, 可为临床预测患者 ALN 转移风险提供参考。

### 关键词

乳腺肿瘤; 青年人; 淋巴转移; 危险因素; 大数据  
中图分类号: R737.9

## Establishment of risk prediction model for axillary lymph node metastasis in breast cancer at early age based on clinicopathologic big data

SONG Wenjing<sup>1</sup>, HE Xin<sup>1</sup>, GONG Pengju<sup>1</sup>, YANG Yan<sup>1</sup>, WEI Lei<sup>2</sup>, ZHANG Jingwei<sup>1</sup>

(1. Department of Breast and Thyroid Surgery, Zhongnan Hospital, Wuhan University, Wuhan 430071, China; 2. Department of Pathology and Pathophysiology, School of Basic Medical Sciences, Wuhan University, Wuhan 430071, China)

收稿日期: 2020-06-04; 修订日期: 2020-10-29。

作者简介: 宋文静, 武汉大学中南医院硕士研究生, 主要从事乳腺癌的临床与基础方面的研究。

通信作者: 张京伟, Email: zjwzhang68@whu.edu.cn

**Abstract**

**Background and Aims:** Over the recent years, the incidence of breast cancer is increasingly shifting to younger population, which is more likely to develop axillary lymph node (ALN) metastasis. Therefore, this study was conducted to determine the influencing factors for ALN metastasis in young breast cancer patients using big-data platform of clinicopathologic information and establish a risk prediction models, so as to provide a reference for the clinical diagnosis and treatment of breast cancer in young adults.

**Methods:** The clinicopathologic data of young patients who were diagnosed with breast cancer and underwent ALN dissection between 2010 and 2015 were selected from the SEER database. The influencing factors for ALN metastasis were determined by univariate and multivariate analysis, and were subsequently visualized by nomogram. The ability of the nomogram to identify patients with different ALN status was quantized using the AUC/C-index. The internal verification of the prediction performance of the nomogram was estimated by bootstrap method (1 000 replicates with a random seed of 12). Furthermore, the data of young patients with newly diagnosed breast cancer from 2015 to 2017 in Zhongnan Hospital of Wuhan University were collected for external validation of the original model.

**Results:** A total of 23 778 young patients with breast cancer was recruited from the SEER database, 39.6% of whom had ALN metastasis. Univariate Logistic regression analysis showed that age, race, location of primary tumor, pathological grade, tumor size, and presence or absence of the chest wall or skin invasion as well as the status of ER, PR and HER-2 were significantly associated with ALN metastasis (all  $P < 0.001$ ). Multivariate Logistic regression analysis showed that age, race, and marital status, laterality, location of primary tumor and grade, tumor size, and presence or absence of the chest wall or skin invasion as well as the status of the ER, PR were independent influencing factors for ALN metastasis (all  $P < 0.05$ ), based on which, the risk prediction model was established. The calibration curve of internal validation indicated a good consistency between the predicted value calculated by the model and the real value (AUC/C-index=0.716). A total of 391 young patients with breast cancer were clinically enrolled as external validation dataset, and 49.9% of them were found to have ALN metastasis at initial diagnosis. The of external validation showed the good predictive ability of the model (AUC/C-index=0.798).

**Conclusion:** The risk prediction model developed using the SEER database for ALN metastasis in young patients with breast cancers has good predictive ability, and it can be used as a reference in clinical practice for estimating ALN metastasis of patients.

**Key words**

Breast Neoplasms; Young Adult; Lymphatic Metastasis; Risk Factors; RBig Data

**CLC number:** R737.9

乳腺癌是对女性威胁最大的恶性肿瘤。据2018年最新癌症报告,乳腺癌仍然是女性发病率和病死率最高的癌症。预计2020年在全球范围内有279 100例新发乳腺癌病例,预计42 690例乳腺癌患者死亡<sup>[1]</sup>。在世界范围内,乳腺癌的发病率呈逐年上升趋势,发病年龄呈年轻化趋势<sup>[2]</sup>。年轻乳腺癌通常侵袭性更强,并具有比老年乳腺癌更差的生存结果<sup>[3]</sup>。腋窝淋巴结(axillary lymph nodes, ALN)状态是乳腺癌患者最重要的预后因素之一<sup>[4-6]</sup>,与无病生存期(disease free survival, DFS)及总生存期(overall survival, OS)密切相关,同时也是决定其手术方式的因素之一。

ALN状态是一个由患者自身情况、肿瘤特征及基因决定的多因素事件<sup>[7-9]</sup>,反映肿瘤进展情况。因此,总结评估患者的临床病理特征对

于指导腋窝手术是非常重要的。为了确定ALN状态,腋窝淋巴结清扫术(axillary lymph node dissection, ALND)曾是浸润性乳腺癌患者的标准手术方式,可使腋窝局部情况获得良好的控制,同时术后的病检结果可以提供准确的肿瘤分期,但术后易出现上肢淋巴水肿,尽管现在乳腺癌根治术后并发患侧上肢水肿可以采用改良静脉淋巴管吻合术,来改善患者的水肿症状<sup>[10]</sup>,但患者需要接受一定的创伤。除此之外,乳腺癌患者还存在乳腺癌根治术后肩关节活动障碍、上臂感觉障碍等并发症,影响患者生活质量。因此,前哨淋巴结活检(sentinel lymph node biopsy, SLNB)已代替ALND成为淋巴结临床阴性(cN0)乳腺癌患者的标准手术方式<sup>[11]</sup>。相比于ALND,SLNB具有创伤小、术后并发症发病率低等优

点<sup>[12-13]</sup>。然而,既往报道接受SLNB的乳腺癌患者有超过70%患者ALN无肿瘤<sup>[14]</sup>。此外,在过去的25年中,乳腺癌的全身和放射疗法的改良已导致区域性复发风险的显著降低。同时,外科医生已经证明,对于ALN疾病负担低的患者,放弃腋窝解剖是安全的。关于肿瘤患者的治疗效果和生活质量的兼顾是临床医生和研究人员面临的共同问题<sup>[15]</sup>。目前,对于ALN的术前评估主要依靠临床触诊、乳腺超声、乳腺钼靶、乳腺增强核磁共振等,影像学检查仍然存在一定的局限性<sup>[16-17]</sup>。近年来,临床风险预测模型在临床实践中得到了广泛的应用,临床医生可以利用现有的患者信息和疾病特征来预测预后不良的风险,并评估患者是否需要接受相应检查与治疗。

本研究旨在通过对SEER数据库中年轻乳腺癌病例进行回顾性研究,筛选与年轻乳腺癌ALN转移相关的因素,并据此构建风险预测模型,协助术前评估年轻乳腺癌患者ALN状态,从而为年轻乳腺癌患者的诊治提供参考。

## 1 资料与方法

### 1.1 对象来源与筛选标准

SEER数据库由美国国家癌症研究所建立,收集有关癌症患者疾病信息和生存结果的资料,包括18个基于人口分布的癌症登记处,覆盖约34.6%的美国人口。使用SEER\*Stat 8.3.5版软件下载并提取2010年1月—2015年12月被诊断为乳腺癌的患者。通过文献检索、既往分析和专家指导,从SEER数据库中检索患者的年龄、种族、性别、婚姻状况、肿瘤位置、边侧、分级、分期、TNM分期、雌激素受体(estrogen receptor, ER)状态、孕激素受体(progesterone receptor, PR)状态、表皮生长因子受体(human epidermal-growth-factor receptor 2, HER-2)状态。研究纳入标准:(1) ICD-O-3仅限于乳腺癌(C500-506; C508-509);(2) 2010—2015年被诊断为乳腺癌的患者;(3) 年龄 $\geq 20$ 岁, <45岁的患者;(4) 单侧、单发肿瘤患者。排除标准:(1) 缺乏种族、婚姻状况、肿瘤位置、边侧、分级、分期、TNM分期、ER状态、PR状态、HER-2状态资料的患者;(2) 原位癌患者;(3) 临床或病理诊断为远处转移的患者;(4) ER和PR状态不详或呈边缘性患者(图1)。

收集2015年1月—2017年12月在武汉大学中南医院被诊断为乳腺癌的病例进行外部验证队列。

本研究经武汉大学中南医院伦理委员会批准。具体纳入标准:(1) 确诊为单侧乳腺癌;(2) 年龄 <45岁;(3) 未进行术前辅助治疗;(4) 行SLNB或ALND;(5) 病理诊断为浸润性乳腺癌;(6) 接受手术时无临床或病理远处转移。

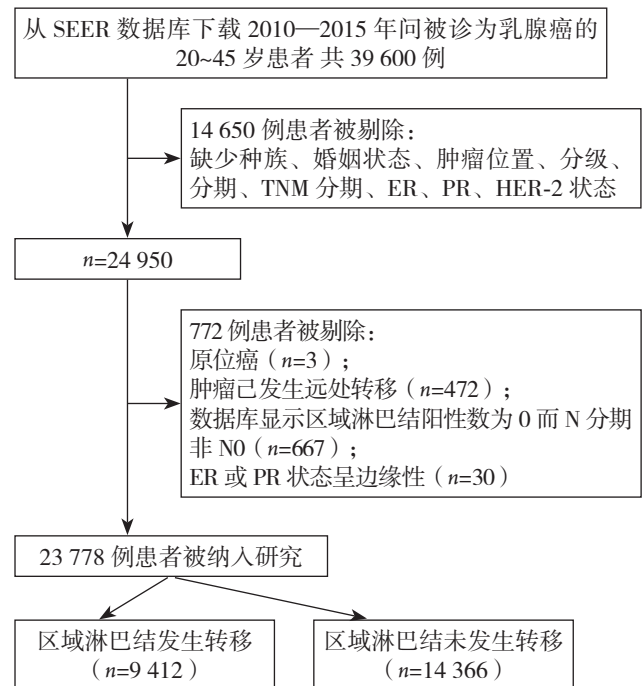


图1 研究对象筛选流程

Figure 1 Case selection process

### 1.2 变量定义和注释

所有病例均按第七届美国癌症联合委员会 TNM 分类法进行重新分类:根据有无胸壁或皮肤侵犯情况,将 T1、T2、T3 期的未见胸壁或皮肤侵犯的肿瘤患者合并为一组, T4 期肿瘤患者单独一组。年龄每增加 5 岁划分一组;种族分为白种人、黑种人和其他人种;肿瘤大小为每增大 1 cm 划分一组, >5 cm 的肿瘤统为一组。重新定义婚姻状况分组:已婚、未婚和离婚(包括分居、离婚、丧偶)。乳腺癌亚型定义: luminal A: HER-2(-)/ER(+)/PR(+); luminal B: HER-2(-)/ER(+)/PR(-); HER-2 阳性: HER-2(+); 三阴性: HER-2(-)/ER(-)/PR(-)。

### 1.3 模型建立与验证

连续变量分组后和分类变量的数据分别按频率和百分比汇总。采用 Logistic 回归进行单因素和多因素回归分析,筛选  $P < 0.05$  的变量确定为 ALN 状态的影响因素。优势比(odds ratio, OR)以 95% 置信区间(confidence interval, CI)表示。

基于Logistic回归建立年轻乳腺癌ALN转移的风险预测模型并通过列线图(nomogram)可视化,通过另一数据集对模型的预测性能和区分度进行测试。通过接收者工作特征曲线(receiver operating characteristic, ROC)曲线下面积(area under the curve, AUC)/C指数(C-index)量化列线图区分不同ALN状态患者的能力,表示模型预测和实际结果的一致性。C指数范围为0~1,1表示完全一致,0表示完全不一致。通常认为,C指数>0.7,模型预测能力较好。采用bootstrap方法(1 000次重复,随机数种子设置为12)进行列线图预测性能验证的评估。

#### 1.4 统计学处理

所有假设检验均取双侧, $P<0.05$ 为差异有统

计学意义。采用SPSS 25.0版本、Graphpad Prism 7.00版本、R软件3.6.1版本进行统计分析及可视化。

## 2 结果

### 2.1 患者的临床病理特征

经筛选共23 778例患者被纳入研究,其中女性23 715例(99.7%),男性63例(0.3%)。原发于左侧的乳腺癌患者11 862例(49.9%),原发于右侧的乳腺癌患者11 916例(50.1%),原发于乳头的乳腺癌患者64例。luminal A型乳腺癌最多,达15 186例(63.9%),HER-2阳性型乳腺癌患者最少,有1 159例(4.9%)。本研究队列的临床病理特征见表1。

表1 23 778例患者的人口学特征和临床病理特征[n(%)]

Table 1 Demographics and clinicopathologic characteristics of the 23 778 patients [n(%)]

特征	区域淋巴结阳性 (n=9 412)	区域淋巴结阴性 (n=14 366)	所有患者 (n=23 778)	特征	区域淋巴结阳性 (n=9 412)	区域淋巴结阴性 (n=14 366)	所有患者 (n=23 778)
年龄(岁)				分期			
20~24	77 (0.8)	76 (0.5)	153 (0.6)	I	713 (7.6)	8 842 (61.5)	9 555 (40.2)
25~29	391 (4.2)	514 (3.6)	905 (3.8)	II	5 309 (56.4)	5 438 (37.9)	10 747 (45.2)
30~34	1 214 (12.9)	1 550 (10.8)	2 764 (11.6)	III	3 390 (36.0)	86 (0.6)	3 476 (14.6)
35~39	2 704 (28.7)	3 469 (24.1)	6 173 (26.0)	T分期			
40~44	5 026 (53.4)	8 757 (61.0)	13 783 (58.0)	T1	3 089 (32.8)	8 842 (61.5)	11 931 (50.1)
种族				T2	4 633 (49.2)	4 895 (34.1)	9 528 (40.1)
黑色人种	1 315 (14.0)	1 786 (12.4)	3 101 (13.0)	T3	1 381 (14.7)	543 (3.8)	1 924 (8.1)
其他人种	1 217 (12.9)	2 039 (14.2)	3 256 (13.7)	T4	309 (3.3)	86 (0.6)	395 (1.7)
白色人种	6 880 (73.1)	10 541 (73.4)	17 421 (73.3)	肿瘤大小(cm)			
性别				≤1	622 (6.6)	3 465 (24.1)	4 087 (17.2)
女	9 385 (99.7)	14 330 (99.7)	23 715 (99.7)	>1~2	2 515 (26.7)	5 391 (37.5)	7 906 (33.2)
男	27 (0.3)	36 (0.3)	63 (0.3)	>2~3	2 657 (28.2)	3 170 (22.1)	5 827 (24.5)
婚姻状态				>3~4	1 353 (14.4)	1 283 (8.9)	2 636 (11.1)
离异	940 (10.0)	1 434 (10.0)	2 374 (10.0)	>4~5	766 (8.2)	488 (3.4)	1 254 (5.3)
已婚	6 089 (64.7)	9 305 (64.8)	15 394 (64.7)	>5	1 499 (15.9)	569 (4.0)	2 068 (8.7)
未婚	2 383 (25.3)	3 627 (25.2)	6 010 (25.3)	N分期			
边侧				N0	0 (0.0)	14 366 (100.0)	14 366 (60.4)
左侧	4 769 (50.7)	7 093 (49.4)	11 862 (49.9)	N1	6 948 (73.8)	0 (0.0)	6 948 (29.2)
右侧	4 643 (49.3)	7 273 (50.6)	11 916 (50.1)	N2	1 675 (17.8)	0 (0.0)	1 675 (7.0)
肿瘤原发部位				N3	789 (8.4)	0 (0.0)	789 (3.2)
腋尾	84 (0.9)	104 (0.8)	188 (0.8)	肿瘤亚型			
中央区	533 (5.6)	483 (3.4)	1 016 (4.3)	luminal A	6 112 (64.9)	9 074 (63.2)	15 186 (63.9)
内下象限	467 (4.9)	954 (6.6)	1 421 (6.0)	luminal B	1 429 (15.2)	2 056 (14.3)	3 485 (14.7)
外下象限	939 (10.0)	1 384 (9.6)	2 323 (9.7)	HER-2 阳性	526 (5.6)	633 (4.4)	1 159 (4.9)
乳头	33 (0.4)	31 (0.2)	64 (0.3)	三阴性	1 345 (14.3)	2 603 (18.1)	3 948 (16.6)
重叠病变	2 453 (26.1)	3 559 (24.8)	6 012 (25.3)	ER			
内上象限	1 028 (10.9)	2 389 (16.6)	3 417 (14.4)	阴性	2 000 (21.2)	3 480 (24.2)	5 480 (23.0)
外上象限	3 875 (41.2)	5 462 (38.0)	9 337 (39.2)	阳性	7 412 (78.8)	10 886 (75.8)	18 298 (77.0)
病理学分级				PR			
I	766 (8.1)	2 464 (17.2)	3 230 (13.6)	阴性	2 694 (28.6)	4 444 (30.9)	7 138 (30.0)
II	3 834 (40.7)	5 418 (37.7)	9 252 (38.9)	阳性	6 718 (71.4)	9 922 (69.1)	16 640 (70.0)
III	4 777 (50.8)	6 436 (44.8)	11 213 (47.2)	HER-2			
IV	35 (0.4)	48 (0.3)	83 (0.3)	阴性	7 457 (79.2)	11 677 (81.3)	19 134 (80.5)
				阳性	1 955 (20.8)	2 689 (18.7)	4 644 (19.5)

## 2.2 ALN 状态的影响因素

为了确定影响淋巴结转移的因素, 对各临床病理信息进行 Logistic 单因素回归分析, 结果显示, 年龄、种族、肿瘤部位、分级、肿瘤大小、胸壁或皮肤是否受侵以及 ER、PR、HER-2 状态与

ALN 转移有关 (均  $P < 0.001$ )。对各临床病理信息进行 Logistic 多因素回归分析显示, 年龄、种族、婚姻状态、边侧、肿瘤部位、分级、肿瘤大小、胸壁或皮肤是否受侵以及 ER 与 PR 状态是 ALN 转移独立影响因素 (均  $P < 0.05$ ) (表 2)。

表 2 23 778 例年轻乳腺癌患者 ALN 状态影响因素的单变量与多变量回归分析

因素	单变量分析		多变量分析		因素	单变量分析		多变量分析	
	OR (95% CI)	P	OR (95% CI)	P		OR (95% CI)	P	OR (95% CI)	P
年龄 (岁)					病理学分级				
20~24	—	—	—	—	I	—	—	—	—
25~29	0.751 (0.533~1.058)	0.102	0.702 (0.485~1.018)	0.062	II	2.276 (2.078~2.493)	<0.001	1.790 (1.623~1.975)	<0.001
30~34	0.773 (0.558~1.071)	0.121	0.735 (0.517~1.046)	0.088	III	2.388 (2.184~2.611)	<0.001	1.759 (1.585~1.951)	<0.001
35~39	0.769 (0.558~1.060)	0.109	0.742 (0.524~1.052)	0.094	IV	2.346 (1.506~3.653)	<0.001	1.804 (1.104~2.949)	<0.001
40~44	0.566 (0.412~0.779)	<0.001	0.628 (0.444~0.888)	0.009	胸壁或皮肤受侵				
种族					否	—	—	—	—
黑色人种	—	—	—	—	是	5.636 (4.432~7.168)	<0.001	3.092 (2.390~4.001)	<0.001
其他人种	0.811 (0.733~0.896)	<0.001	0.800 (0.716~0.894)	<0.001	肿瘤大小 (cm)				
白色人种	0.886 (0.820~0.958)	0.002	0.904 (0.829~0.985)	0.021	≤ 1	0.068 (0.060~0.078)	<0.001	0.073 (0.064~0.083)	<0.001
性别					> 1~2	0.177 (0.159~0.197)	<0.001	0.185 (0.165~0.206)	<0.001
女	—	—	—	—	> 2~3	0.318 (0.285~0.355)	<0.001	0.332 (0.297~0.372)	<0.001
男	1.145 (0.695~1.887)	0.595	0.722 (0.426~1.222)	0.224	> 3~4	0.400 (0.354~0.453)	<0.001	0.428 (0.377~0.485)	<0.001
婚姻状态					> 4~5	0.596 (0.513~0.692)	<0.001	0.621 (0.533~0.723)	<0.001
离异	—	—	—	—	> 5	—	—	—	—
已婚	0.998 (0.914~1.091)	0.969	1.029 (0.935~1.132)	0.559	肿瘤亚型				
未婚	1.002 (0.910~1.105)	0.963	0.878 (0.790~0.977)	0.017	luminal A	—	—	—	—
边侧					luminal B	1.032 (0.957~1.112)	0.412	—	—
左侧	—	—	—	—	HER-2 阳性	1.234 (1.094~1.391)	0.001	—	—
右侧	0.949 (0.901~1.000)	0.051	0.934 (0.883~0.988)	0.017	三阴性	0.767 (0.713~0.826)	<0.001	—	—
肿瘤原发部位					ER				
腋尾	—	—	—	—	阴性	—	—	—	—
中央区	1.366 (0.999~1.868)	0.051	1.070 (0.767~1.492)	0.692	阳性	1.185 (1.113~1.261)	<0.001	1.405 (1.269~1.556)	<0.001
内下象限	0.606 (0.445~0.825)	0.001	0.515 (0.372~0.713)	<0.001	PR				
外下象限	0.840 (0.623~1.133)	0.253	0.708 (0.516~0.971)	0.032	阴性	—	—	—	—
乳头	1.318 (0.747~2.327)	0.341	1.298 (0.701~2.401)	0.407	阳性	1.117 (1.055~1.182)	<0.001	1.297 (1.179~1.425)	<0.001
重叠区域	0.853 (0.637~1.143)	0.287	0.656 (0.482~0.894)	0.007	HER-2				
内上象限	0.533 (0.396~0.717)	<0.001	0.421 (0.308~0.576)	<0.001	阴性	—	—	—	—
外上象限	0.878 (0.657~1.174)	0.381	0.703 (0.517~0.956)	0.024	阳性	1.138 (1.067~1.215)	<0.001	1.024 (0.953~1.100)	0.516

## 2.3 模型的建立

基于多因素 Logistic 回归确定预测因素 (年龄、种族、婚姻状况、边侧、原发肿瘤位置、胸壁或皮肤是否受侵、肿瘤大小、分级、ER 和 PR 状态) 建立预测年轻乳腺癌淋巴结转移的模型并以列线图可视化 (图 2)。从列线图上可以看出, 种族、婚姻状况、边侧对淋巴结状态的影响较小, 而肿瘤大小、位置、胸壁或皮肤是否受侵、病理分级对 ALN 状态的影响较大。值得注意的是相比于病理学 I 级肿瘤, II、III、IV 级肿瘤具有更高的 ALN 转移风险 ( $OR = 1.790$ 、 $1.759$ 、 $1.804$ , 均  $P < 0.001$ ),

但在 II、III、IV 级这 3 组间未见明显差异。

## 2.4 模型的内部验证

采用 bootstrap 方法对获得的列线图进行内部验证, AUC 为 0.716 (95% CI = 0.710~0.723), 校准曲线显示预测值与真实值之间存在良好的一致性, 显示了该模型的良好精度 (图 3)。

## 2.5 外部验证数据集

收集了来自医院的 391 例年轻乳腺癌患者作为外部验证队列, 这些病例的临床病理特征如表 3 所示。术前乳腺肿瘤活检与术后病检的分级、PR、ER、HER-2 结果无差异。该组年轻乳腺癌患者

ALN转移发生率49.9%。

### 2.6 模型的外部验证

外部验证得到的AUC为0.798 (95% CI=0.752-

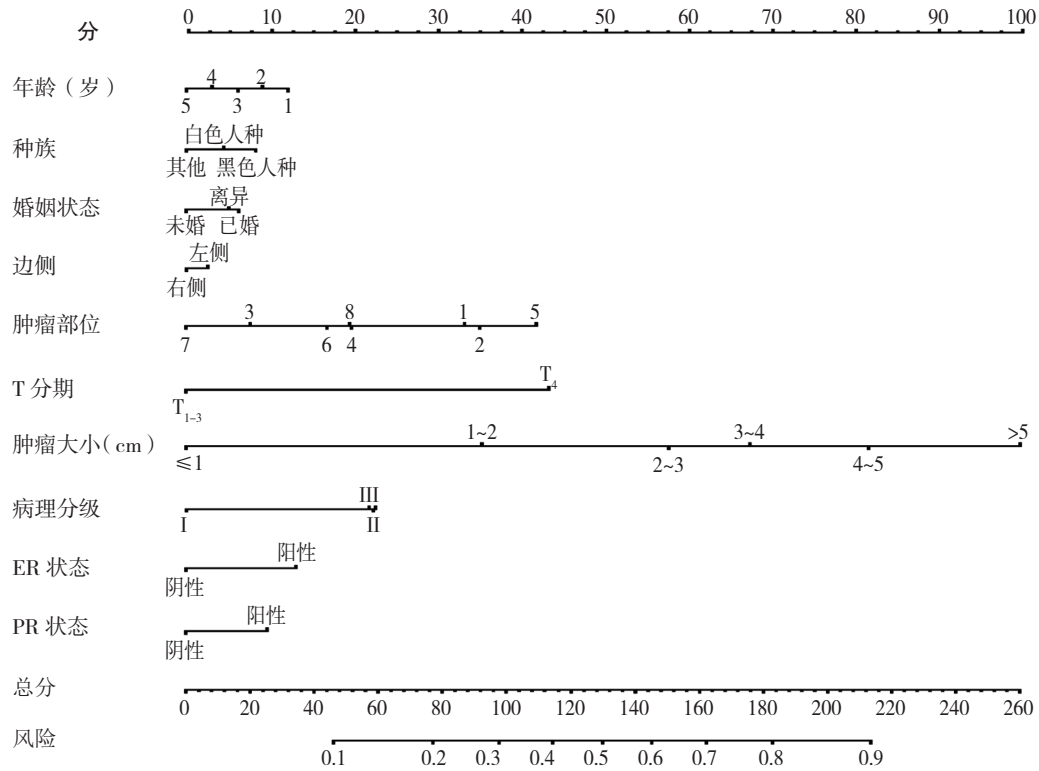


图 2 预测年轻乳腺癌 ALN 转移的列线图 (年龄: 1=20~24 岁, 2=25~29 岁, 3=30~34 岁, 4=35~39 岁, 5=40~44 岁; 肿瘤部位: 1= 腋尾部, 2= 乳腺中央, 3= 内下象限, 4= 外下象限, 5= 乳头, 6= 重叠区域, 7= 内上象限, 8= 外上象限; T1~3: 胸壁或皮肤未见肿瘤侵犯, T4: 胸壁或皮肤可见肿瘤侵犯)

Figure 2 Nomogram for predicting ALN metastasis in young breast cancer patients (Age: 1=20–24 years old, 2=25–29 years old, 3=30–34 years old, 4=35–39 years old, 5=40–44 years old; Primary Site: 1=axillary tail, 2=central portion of breast, 3=lower-inner quadrant of breast, 4=lower-outer quadrant of breast, 5=nipple, 6=overlapping lesion of breast, 7=upper-inner quadrant of breast, 8=upper-outer quadrant of breast; T1–3: absence of tumor invasion of chest wall or skin, T4: presence of invasion of chest wall or skin)

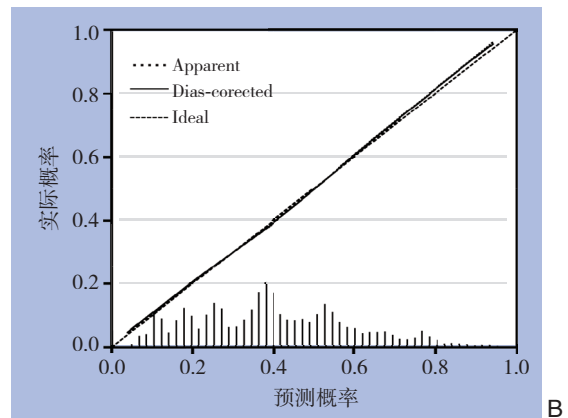
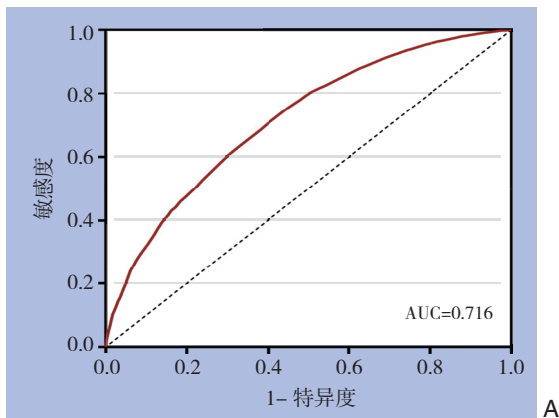


图 3 模型的内部验证 A: ROC 曲线; B: 校准曲线

Figure 3 Internal validation of the model A: ROC curve; B: Calibration plot

表 3 391 例患者的人口学特征及临床病理特征 [n (%)]

Table 3 Demographics and clinicopathologic characteristics of the 391 patients [n (%)]

特征	区域淋巴结阳性 (n=195)	区域淋巴结阴性 (n=196)	所有患者 (n=391)	特征	区域淋巴结阳性 (n=195)	区域淋巴结阴性 (n=196)	所有患者 (n=391)
年龄 (岁)				分级			
20~24	0 (0.0)	0 (0.0)	0 (0.0)	I	6 (3.1)	16 (8.2)	22 (5.6)
25~29	0 (0.0)	4 (2.0)	4 (1.0)	II	91 (46.6)	100 (51.0)	191 (48.8)
30~34	39 (20.0)	27 (13.8)	66 (16.9)	III	98 (50.3)	80 (40.8)	178 (45.6)
35~39	42 (21.5)	26 (13.3)	68 (17.4)	IV	0 (0.0)	0 (0.0)	0 (0.0)
40~44	114 (58.5)	139 (70.9)	253 (64.7)	T 分期			
种族				T1	67 (34.4)	95 (48.5)	162 (41.4)
黑色人种	0 (0.0)	0 (0.0)	0 (0.0)	T2	48 (24.6)	74 (37.8)	122 (31.2)
其他人种	195 (100.0)	196 (100.0)	391 (100.0)	T3	52 (26.7)	26 (13.2)	78 (20.0)
白色人种	0 (0.0)	0 (0.0)	0 (0.0)	T4	28 (14.4)	1 (0.5)	29 (7.4)
性别				肿瘤大小 (cm)			
女	195 (100.0)	196 (100.0)	391 (100.0)	≤ 1	8 (4.1)	24 (12.2)	32 (8.2)
男	0 (0.0)	0 (0.0)	0 (0.0)	> 1~2	21 (10.8)	77 (39.3)	98 (25.1)
婚姻状态				> 2~3	92 (47.2)	65 (33.2)	157 (40.1)
离异	0 (0.0)	0 (0.0)	0 (0.0)	> 3~4	39 (20.0)	14 (7.2)	53 (13.5)
已婚	195 (100.0)	196 (100.0)	391 (100.0)	> 4~5	17 (8.7)	4 (2.0)	21 (5.4)
未婚	0 (0.0)	0 (0.0)	0 (0.0)	> 5	18 (9.2)	12 (6.1)	30 (7.7)
边侧				N 分期			
左侧	100 (51.3)	95 (48.5)	195 (49.9)	N0	0 (0.0)	196 (100.0)	196 (50.1)
右侧	95 (48.7)	101 (51.5)	196 (50.1)	N1	103 (52.8)	0 (0.0)	103 (26.3)
肿瘤原发部位				N2	65 (33.3)	0 (0.0)	65 (16.7)
腋尾	3 (1.5)	0 (0.0)	3 (0.8)	N3	27 (13.8)	0 (0.0)	27 (6.9)
中央区	4 (2.1)	2 (1.0)	6 (1.5)	ER			
内下象限	13 (6.7)	36 (18.4)	49 (12.5)	阴性	37 (19.0)	63 (32.1)	100 (25.6)
外下象限	33 (16.9)	16 (8.2)	49 (12.5)	阳性	158 (81.0)	133 (67.9)	291 (74.4)
乳头	0 (0.0)	0 (0.0)	0 (0.0)	PR			
重叠区域	0 (0.0)	0 (0.0)	0 (0.0)	阴性	61 (31.3)	72 (36.7)	133 (34.0)
内上象限	12 (6.2)	83 (42.3)	95 (24.3)	阳性	134 (68.7)	124 (63.3)	258 (66.0)
外上象限	130 (66.6)	59 (30.1)	189 (48.4)	HER-2			
				阴性	119 (61.0)	126 (64.3)	245 (62.7)
				阳性	76 (39.0)	70 (35.7)	146 (37.3)

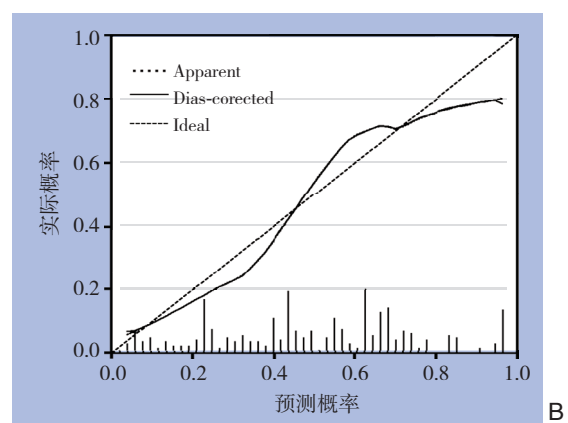
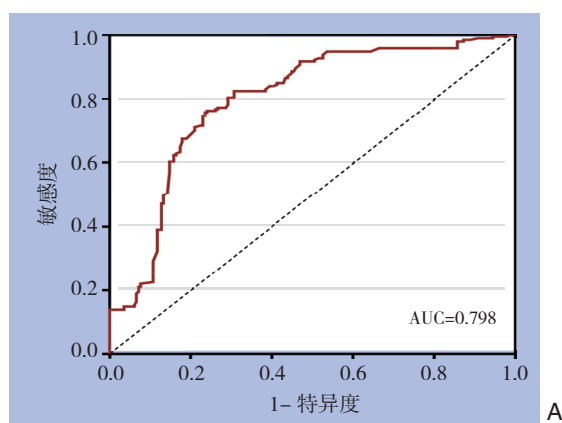


图 4 模型的外部验证 A: ROC 曲线; B: 校准曲线

Figure 4 External validation of the model A: ROC curve; B: Calibration plot

0.844), 其ROC曲线与校准曲线如图4A、4B所示, 说明该模型对外部数据集应用的可行性。

### 3 讨论

在临床实践中, 年轻的乳腺癌具有特殊的临床病理的特点, 如分期晚, 分级高, 淋巴结阳性率高, 淋巴结转移数量较多、远处转移率高, 容

易合并血管内瘤栓, Ki-67高, HER-2阳性率高和雌激素受体阳性率较低等, 并更可能存在家族性易感基因<sup>[18-19]</sup>。年轻患者对生活质量有更高的要求, 因此, 如何在保证治疗效果的同时降低创伤, 如何更早、更准确地预测和评估乳腺癌的严重程度, 进而给予个性化治疗是肿瘤领域的研究热点。预测ALN状态在年轻女性乳腺癌手术方案的选择上起重要作用, 可以帮助临床医生确定哪些患者可以从治疗中获益。因此, 基于SEER数据库, 本研究通过多因素回归分析发现年龄、种族、婚姻状况、边侧、原发肿瘤位置、胸壁或皮肤是否受侵、肿瘤大小、分级、ER和PR状态是影响年轻乳腺癌患者ALN转移的独立危险因素, 并据此建立了一个的年轻乳腺癌患者ALN状态的风险预测模型。

年龄被认为是乳腺癌患者淋巴结状态的独立预测因素。既往对于年轻乳腺癌患者年龄与ALN状态的关系的研究较少, 全年龄段乳腺癌患者的分析结果显示, 年轻乳腺癌通常侵袭性更强, ALN转移率更高, 并具有比老年乳腺癌更差的生存结果。比利时Leuven大学医院的一项回顾性研究<sup>[20]</sup>发现, 年龄对淋巴结受累的影响是分段的。年龄与ALN状态之间的相关性曲线呈非线性的“U”型, 在70岁之前, 每增加10岁, 淋巴结转移的风险降低13% ( $OR=0.87$ ,  $95\% CI=0.78\sim 0.96$ ,  $P<0.001$ ), 但70岁以后, 年龄与淋巴结转移风险呈正相关。Botteri等<sup>[21]</sup>的研究也发现了类似的相关性曲线, 但转折点在65岁。在本研究中, 年轻乳腺癌患者( $<45$ 岁)的淋巴结转移率为39.6%; 年龄每增加5岁, 淋巴结转移风险降低13.9% ( $OR=0.861$ ,  $95\% CI=0.836\sim 0.887$ ,  $P<0.001$ )。通过分析结果可以发现, 年轻患者一旦被诊断为乳腺癌, 应早期积极采取措施进行治疗, 以期达到较好的预后和生活质量。Zhang等<sup>[22]</sup>在回顾性分析1 671例早期乳腺癌患者淋巴结转移的风险因素时发现乳头下方和乳腺外侧肿瘤更容易发生ALN转移。Janni等<sup>[23]</sup>回顾性分析的结果也表明, ALN转移与肿瘤位置显著相关, 肿瘤位于外侧的病例的平均ALN转移数量增加了29%, 然而, Malter等<sup>[24]</sup>在整个乳腺癌人群中进行多因素分析中发现肿瘤部位并不是淋巴结状态的独立影响因素。在本研究的单因素和多因素Logistic分析中, 肿瘤位置被认为与淋巴结状态有关 ( $P<0.001$ ), 位于乳头和中央部位的乳腺癌

最容易发生ALN转移, 但未发现任意两个不同的肿瘤部位的淋巴结转移风险存在显著性差异。临床医生应警惕乳头和中央区乳腺癌患者在进行中央区保乳或全乳房切除时的ALN转移。肿瘤大小每增加1 cm分为一组。单因素和多因素Logistic分析显示, 肿瘤大小是影响ALN状态的因素之一 ( $P<0.001$ ), 肿瘤每增大1 cm, 淋巴结转移率增加53.4% ( $OR=1.534$ ,  $95\% CI=1.507\sim 1.563$ ,  $P<0.001$ )。张彦收等<sup>[25]</sup>收集了196例前哨淋巴结宏转移临床早期乳腺癌患者临床资料, 分析发现原发肿瘤大小、阳性前哨淋巴结个数是前哨淋巴结宏转移的临床早期乳腺癌患者前哨淋巴结转移重要因素。Sopik等<sup>[26]</sup>分析发现对于大约7~60 mm大小的乳腺癌, 肿瘤大小与淋巴结转移风险之间存在明显的线性相关性, 与本研究的分析结果一致。既往研究<sup>[27-31]</sup>表明, ER、PR和HER-2状态是淋巴结状态的预测指标, 谢竞等<sup>[32]</sup>检测了84例青年乳腺癌肿瘤样本中的C-erbB-2蛋白表达, 结果显示C-erbB-2蛋白在青年乳腺癌中的阳性表达水平与ALN的转移存在正相关。而本项研究的多因素分析结果显示, HER-2状态对年轻乳腺癌淋巴结转移风险未见显著影响 ( $P=0.516$ )。结果的差异可能与纳入研究人群的年龄、分析的影响因素、采取的统计方法不同有关。

由于外部验证可以进一步测试风险预测模型的潜在适用性, 本研究使用武汉大学中南医院的391例病例对模型进行了测试。内外和外部验证均表明预测模型具有较好的预测价值。虽然列线图在预测ALN转移方面显示出了良好的准确性, 但是在解释分析结果时必须考虑到数据的局限性。首先, 使用回顾性研究的数据导致了选择偏倚的可能性。例如, 一些患者可能已经接受了术前新辅助治疗, 因此手术中获得的信息(如TNM分期)可能不如初诊时准确, 这可能导致淋巴结转移率比实际上低。其次, 基于SEER数据库建立模型时, 肿瘤的大小和T分期均来自于肿瘤切除后的数据, 而我们在术前进行风险预测时, 只能依赖临床体检、乳腺超声、X线、MRI影像学检查。此外, 多因素Logistic回归分析显示, 年轻乳腺癌患者种族和肿瘤边侧对淋巴结状态也有明显影响。然而, 根据列线图, 肿瘤和边侧项目的风险得分低于10分, 对风险总分的贡献并不大。此外, 既往研究未发现淋巴结状态与种族或边侧的关系, 需要联合更多的区域医疗中心, 纳入更多的数据



资料以供进一步研究。尽管有这些局限性, 列线图的内部和外部验证显示AUC大于0.7, 显示了较好的预测能力, 希望能对临床工作提供一定的借鉴。

总而言之, 本研究基于人群的队列建立了年轻乳腺癌患者的ALN状态的风险预测模型, 该模型应用现有的临床病理因素来评估ALN转移的风险, 并且显示出较好的预测能力。大数据的分析发展有望进一步提高肿瘤的准确诊断和治疗, 为人类医学提供更好的帮助和支持。

### 参考文献

- [1] Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020[J]. *CA Cancer J Clin*, 2020, 70(1):7–30. doi: 10.3322/caac.21590.
- [2] Esteva FJ, Hubbard-Lucey VM, Tang J, et al. Immunotherapy and targeted therapy combinations in metastatic breast cancer[J]. *Lancet Oncol*, 2019, 20(3):e175–186. doi: 10.1016/S1470-2045(19)30026-9.
- [3] Kataoka A, Iwamoto T, Tokunaga E, et al. Young adult breast cancer patients have a poor prognosis independent of prognostic clinicopathological factors: a study from the Japanese Breast Cancer Registry[J]. *Breast Cancer Res Treat*, 2016, 160(1):163–172. doi: 10.1007/s10549-016-3984-8.
- [4] Capanema Saliba Franco R, Salvador Silva HM, Faria Almeida LL, et al. Sentinel lymph node prognosis value in breast cancer[J]. *Breast J*, 2019, 25(3):557–559. doi: 10.1111/tbj.13271.
- [5] Liu D, Chen Y, Deng M, et al. Lymph node ratio and breast cancer prognosis: a meta-analysis[J]. *Breast Cancer*, 2014, 21(1):1–9. doi: 10.1007/s12282-013-0497-8.
- [6] Andersson Y, Bergkvist L, Frisell J, et al. Long-term breast cancer survival in relation to the metastatic tumor burden in axillary lymph nodes[J]. *Breast Cancer Res Treat*, 2018, 171(2):359–369. doi: 10.1007/s10549-018-4820-0.
- [7] Sawaki M, Idota A, Ichikawa M, et al. Impact of intrinsic subtype on predicting axillary lymph node metastasis in breast cancer[J]. *Oncol Lett*, 2014, 8(4):1707–1712. doi: 10.3892/ol.2014.2333.
- [8] Smeets A, Ryckx A, Belmans A, et al. Impact of tumor chronology and tumor biology on lymph node metastasis in breast cancer[J]. *Springerplus*, 2013, 2:480. doi: 10.1186/2193-1801-2-480.
- [9] Li K, Li GD, Sun LY, et al. PTEN and SHIP: Impact on lymphatic metastasis in breast cancer[J]. *J Cancer Res Ther*, 2018, 14(Supplement):S937–941. doi: 10.4103/0973-1482.193894.
- [10] 武政喜, 张阿龙, 张志业. 淋巴管吻合术治疗乳腺癌术后上肢水肿: 附120例报告[J]. *中国普通外科杂志*, 2016, 25(11):1671–1674. doi:10.3978/j.issn.1005-6947.2016.11.026.
- [11] Wu ZX, Zhang AL, Zhang ZY. Lymphatic vessel anastomosis in treatment of upper limb edema after breast cancer surgery: a report of 120 cases[J]. *Chinese Journal of General Surgery*, 2016, 25(11):1671–1674. doi:10.3978/j.issn.1005-6947.2016.11.026.
- [12] Manca G, Rubello D, Tardelli E, et al. Sentinel Lymph Node Biopsy in Breast Cancer: Indications, Contraindications, and Controversies[J]. *Clin Nucl Med*, 2016, 41(2):126–133. doi: 10.1097/RLU.0000000000000985.
- [13] Zahoor S, Haji A, Battoo A, et al. Sentinel Lymph Node Biopsy in Breast Cancer: A Clinical Review and Update[J]. *J Breast Cancer*, 2017, 20(3):217–227. doi: 10.4048/jbc.2017.20.3.217.
- [14] Manca G, Tardelli E, Rubello D, et al. Sentinel lymph node biopsy in breast cancer: a technical and clinical appraisal[J]. *Nucl Med Commun*, 2016, 37(6):570–576. doi: 10.1097/MNM.0000000000000489.
- [15] Qiu SQ, Zhang GJ, Jansen L, et al. Evolution in sentinel lymph node biopsy in breast cancer[J]. *Crit Rev Oncol Hematol*, 2018, 123:83–94. doi: 10.1016/j.critrevonc.2017.09.010.
- [16] Sener SF. Advances in axillary surgery for breast cancer 2019[J]. *J Surg Oncol*, 2020, 121(1):20–24. doi: 10.1002/jso.25591.
- [17] Choi HY, Park M, Seo M, et al. Preoperative Axillary Lymph Node Evaluation in Breast Cancer: Current Issues and Literature Review[J]. *Ultrasound Q*, 2017, 33(1):6–14. doi: 10.1097/RUQ.0000000000000277.
- [18] 陈洁文, 李姣, 马微妹, 等. 早期浸润性乳腺癌腋窝淋巴结转移与术前MRI及临床病理特征的相关性分析[J]. *中国临床医学影像杂志*, 2020, 31(9):618–623. doi:10.12117/jccmi.2020.09.003.
- [19] Chen JW, Li J, Ma WM, et al. Correlation analysis between axillary lymph node metastasis and preoperative MRI as well as clinicopathological features in early invasive breast cancer[J]. *Journal of China Clinic Medical Imaging*, 2020, 31(9):618–623. doi:10.12117/jccmi.2020.09.003.
- [20] Lee HB, Han W. Unique features of young age breast cancer and its management[J]. *J Breast Cancer*, 2014, 17(4):301–307. doi: 10.4048/jbc.2014.17.4.301.
- [21] Johnson RH, Anders CK, Litton JK, et al. Breast cancer in adolescents and young adults[J]. *Pediatr Blood Cancer*, 2018, 65(12):e27397. doi: 10.1002/pbc.27397.
- [22] Wildiers H, Van Calster B, van de Poll-Franse LV, et al. Relationship between age and axillary lymph node involvement in women with breast cancer[J]. *J Clin Oncol*, 2009, 27(18):2931–2937. doi: 10.1200/JCO.2008.16.7619.
- [23] Botteri E, Bagnardi V, Goldhirsch A, et al. Axillary lymph node involvement in women with breast cancer: does it depend on age?[J]. *Clin Breast Cancer*, 2010, 10(4):318–321. doi: 10.3816/

- CBC.2010.n.042.
- [22] Zhang Y, Li J, Fan Y, et al. Risk factors for axillary lymph node metastases in clinical stage T1–2N0M0 breast cancer patients[J]. *Medicine (Baltimore)*, 2019, 98(40):e17481. doi: 10.1097/MD.00000000000017481.
- [23] Janni W, Rack B, Sommer H, et al. Intra-mammary tumor location does not influence prognosis but influences the prevalence of axillary lymph-node metastases[J]. *J Cancer Res Clin Oncol*, 2003, 129(9):503–510. doi: 10.1007/s00432–003–0465–3.
- [24] Malter W, Hellmich M, Badian M, et al. Factors Predictive of Sentinel Lymph Node Involvement in Primary Breast Cancer[J]. *Anticancer Res*, 2018, 38(6):3657–3662. doi: 10.21873/anticancer.12642.
- [25] 张彦收, 刘运江. 前哨淋巴结宏转移的早期乳腺癌非前哨淋巴结转移危险因素分析[J]. *中国普通外科杂志*, 2019, 28(5):606–611. doi:10.7659/j.issn.1005–6947.2019.05.013.
- Zhang YS, Liu YJ. Analysis of risk factors for non-sentinel lymph node metastasis in early breast cancer with sentinel lymph node macrometastasis[J]. *Chinese Journal of General Surgery*, 2019, 28(5):606–611. doi:10.7659/j.issn.1005–6947.2019.05.013.
- [26] Sopik V, Narod SA. The relationship between tumour size, nodal status and distant metastases: on the origins of breast cancer[J]. *Breast Cancer Res Treat*, 2018, 170(3):647–656. doi: 10.1007/s10549–018–4796–9.
- [27] Öz B, Akcan A, Doğan S, et al. Prediction of nonsentinel lymph node metastasis in breast cancer patients with one or two positive sentinel lymph nodes[J]. *Asian J Surg*, 2018, 41(1):12–19. doi: 10.1016/j.asjsur.2016.06.001.
- [28] Crabb SJ, Cheang MC, Leung S, et al. Basal breast cancer molecular subtype predicts for lower incidence of axillary lymph node metastases in primary breast cancer[J]. *Clin Breast Cancer*, 2008, 8(3):249–256. doi: 10.3816/CBC.2008.n.028.
- [29] Wiechmann L, Sampson M, Stempel M, et al. Presenting features of breast cancer differ by molecular subtype[J]. *Ann Surg Oncol*, 2009, 16(10):2705–2710. doi: 10.1245/s10434–009–0606–2.
- [30] Ugras S, Stempel M, Patil S, et al. Estrogen receptor, progesterone receptor, and HER2 status predict lymphovascular invasion and lymph node involvement[J]. *Ann Surg Oncol*, 2014, 21(12):3780–3786. doi: 10.1245/s10434–014–3851–y.
- [31] 龙丞, 罗铭, 曾健. 临床腋窝淋巴结阴性乳腺癌前哨淋巴结转移 logistic 回归模型的建立和验证[J]. *广东医学*, 2019, 40(24):3388–3394. doi:10.13820/j.cnki.gdyx.20192101.
- Long C, Luo M, Zeng J. Logistic regression model of sentinel lymph node metastasis in clinical axillary node-negative breast cancer[J]. *Guangdong Medical Journal*, 2019, 40(24):3388–3394. doi:10.13820/j.cnki.gdyx.20192101.
- [32] 谢竞, 赵晓燕, 赵冰, 等. C-erbB-2 蛋白表达与青年乳腺癌腋窝淋巴结转移相关性及其生物学意义[J]. *实用癌症杂志*, 2019, 34(11):1776–1779. doi:10.3969/j.issn.1001–5930.2019.11.011.
- Xie J, Zhao XY, Zhao B, et al. Correlation between C-erbB-2 Protein Expression and Axillary Lymph Node Metastasis in Young Breast Cancer and Its Biological Significance[J]. *The Practical Journal of Cancer*, 2019, 34(11):1776–1779. doi:10.3969/j.issn.1001–5930.2019.11.011.

( 本文编辑 宋涛 )

本文引用格式: 宋文静, 贺鑫, 龚鹏举, 等. 基于临床病理大数据的早发乳腺癌腋窝淋巴结转移风险预测模型构建[J]. *中国普通外科杂志*, 2020, 29(11):1293–1302. doi:10.7659/j.issn.1005–6947.2020.11.002

Cite this article as: Song WJ, He X, Gong PJ, et al. Establishment of risk prediction model for axillary lymph node metastasis in breast cancer at early age based on clinicopathologic big data[J]. *Chin J Gen Surg*, 2020, 29(11):1293–1302. doi:10.7659/j.issn.1005–6947.2020.11.002