

Multicenter study of the clinicopathological features and recurrence risk prediction model of early-stage breast cancer with low-positive human epidermal growth factor receptor 2 expression in China (Chinese Society of Breast Surgery 021)

Ling Xin¹, Qian Wu¹, Chongming Zhan², Hongyan Qin², Hongyu Xiang¹, Ling Xu¹, Jingming Ye¹, Xuening Duan¹, Yinhua Liu¹, Chinese Society of Breast Surgery (CSBrS), Chinese Society of Surgery of Chinese Medical Association

¹Breast Disease Center, Peking University First Hospital, Beijing 100034, China;

²Jun Yi Tai He Medical Technology Co. Ltd, Beijing 100024, China.

Abstract

Background : Breast cancer with low-positive human epidermal growth factor receptor 2 (HER2) expression has triggered further refinement of evaluation criteria for HER2 expression. We studied the clinicopathological features of early-stage breast cancer with low-positive HER2 expression in China and analyzed prognostic factors.

Methods : Clinical and pathological data and prognostic information of patients with early-stage breast cancer with low-positive HER2 expression treated by the member units of the Chinese Society of Breast Surgery and Chinese Society of Surgery of Chinese Medical Association, from January 2015 to December 2016 were collected. The prognostic factors of these patients were analyzed.

Results : Twenty-nine hospitals provided valid cases. From 2015 to 2016, a total of 25,096 cases of early-stage breast cancer were treated, 7642 (30.5%) of which had low-positive HER2 expression and were included in the study. After ineligible cases were excluded, 6486 patients were included in the study. The median follow-up time was 57 months (4–76 months). The disease-free survival rate was 92.1% at 5 years, and the overall survival rate was 97.4% at 5 years. At the follow-up, 506 (7.8%) cases of metastasis and 167 (2.6%) deaths were noted. Multivariate Cox regression analysis showed that tumor stage, lymphovascular invasion, and the Ki67 index were related to recurrence and metastasis ($P < 0.05$). The recurrence risk prediction model was established using a machine learning model and showed that the area under the receiving operator characteristic curve was 0.815 (95% confidence interval: 0.750–0.880).

Conclusions : Early-stage breast cancer patients with low-positive HER2 expression account for 30.5% of all patients. Tumor stage, lymphovascular invasion, and the Ki67 index are factors affecting prognosis. The recurrence prediction model for breast cancer with low-positive HER2 expression based on a machine learning model had a good clinical reference value for predicting the recurrence risk at 5 years.

Trial registration : ChiCTR.org.cn, ChiCTR2100046766.

Keywords: Breast tumor; Low-positive HER2 expression; Multicenter; CSBrS research; Recurrence risk prediction model

Introduction

In the 21st century, breast cancer has entered a new era of classified treatment. The successful development of targeted drugs, such as trastuzumab and pertuzumab, has had significant survival benefits for patients with human epidermal growth factor receptor 2 (HER2)-positive breast cancer.^[1,2] The successful development of antibody-drug conjugates (ADCs), especially the finding

that DS-8201a has continuous disease control in advanced breast cancer with low-positive HER2 expression,^[3] has triggered further refinement of evaluation criteria for HER2 expression status and in-depth consideration of clinical issues related to the prognosis of breast cancer with low-positive HER2 expression status.

To analyze the clinical and pathological characteristics of early-stage breast cancer with low-positive HER2

Access this article online

Quick Response Code:



Website:
www.cmj.org

DOI:
10.1097/CM9.0000000000002056

Correspondence to: Prof. Yinhua Liu, Breast Disease Center, Peking University First Hospital, Beijing 100034, China
E-Mail: liuyinhua7520@163.com

Copyright © 2022 The Chinese Medical Association, produced by Wolters Kluwer, Inc. under the CC-BY-NC-ND license. This is an open access article distributed under the terms of the Creative Commons Attribution-Non Commercial-No Derivatives License 4.0 (CCBY-NC-ND), where it is permissible to download and share the work provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal.

Chinese Medical Journal 2022;135(6)

Received: 13-11-2021; Online: 07-03-2022 Edited by: Yanjie Yin and Xiuyuan Hao

expression in China and related factors affecting prognosis, the Chinese Society of Breast Surgery (CSBrS) organized a multicenter study (CSBrS-021). The clinical and pathological data of 25,096 early-stage breast cancer patients admitted to 29 hospitals of the CSBrS from January 2015 to December 2016 were analyzed and studied, and a machine learning model was used to establish a metastasis risk prediction model to predict recurrence and metastasis events. The findings are reported below.

Methods

Ethical approval

This study was approved by the Ethics Committee of Peking University First Hospital, and completed licensing filing with China Human Genetic Resources.

Study subjects

Patients diagnosed with early invasive breast cancer at CSBrS hospitals from January 1, 2015 to December 31, 2016 were selected as research subjects. To ensure the consistency of the HER2 diagnostic criteria applied to the included patients, the HER2 status was evaluated in accordance with the American Society of Clinical Oncology/College of American Pathologists (ASCO/CAP)^[4] 2013 HER2 diagnostic criteria for breast cancer.

Participating institutions

The CSBrS includes 40 tertiary grade A class hospitals in China as member institutions, all of which meet the qualifications of having an independent breast surgery ward and a pathology laboratory that independently issues pathological reports for breast cancer tissues. A total of 29 CSBrS member hospitals participated in the study and provided qualified cases.

Inclusion criteria

The inclusion criteria are as follows: females with invasive breast cancer confirmed by pathological biopsy; cases with no distant metastasis found on clinical examination; HER2 test results consistent with a diagnosis of low-positive HER2 expression, that is, immunohistochemistry (IHC) 1 +, or IHC 2 + with negative *in situ* hybridization (ISH); previous R0 mastectomy; systematic treatment with the recommended regimen completed according to guidelines^[5]; and complete tumor IHC examination and followup information.

Exclusion criteria

Exclusion criteria are as follows: males with breast cancer, cases of first diagnosis of stage IV breast cancer or metastatic breast cancer or bilateral breast cancer; HER2 positivity or IHC 0 status; an inability to undergo standard systemic treatment and surgical treatment; a history of previous tumor treatment; and incomplete information on tumor IHC and follow-up data.

HER2 status evaluation

Laboratory qualifications

In this study, the pathology laboratory of the reference hospital was required to have a laboratory qualification certificate for performing HER2 detection of breast cancer. The pathology laboratory was required to be certified by the National Pathology Quality Control Center (PQCC) or the International Organization for Standardization/China National Accreditation Service for Conformity Assessment (ISO15189/CNAS-CL02) and to conduct HER2 detection according to sound laboratory standard operating procedures to ensure the reliability and accuracy of the detection results.

Reagent used for HER2 detection

The pathology laboratory of the participating hospital was required to use detection kits certified by the National Medical Products Administration (NMPA) for HER2 detection of breast cancer: the acceptable HER2 probe kits using IHC included VENTANA (4B5, Roche, USA), HercepTestTM (Dako, Denmark), and domestic HER2 detection kits approved by the NMPA (IHC). The acceptable fluorescence *in situ* hybridization (FISH) kits included PathVysion (Abbott molecular, USA), Histra (Jokoh, Japan), Inform (Roche, USA), ZytoLight (Zyto-Vision GmbH, Germany), and CFDA-approved domestic HER2 gene amplification test kits (FISH).

HER2 test standard

The HER2 testing and interpretation standards followed the HER2 diagnostic guidelines revised by ASCO/CAP in 2013.^[4] HER2 positive standard: IHC 3 + or IHC 2 + and ISH positive; low-positive HER2 expression standard: IHC 1 + or IHC 2 + and ISH negative; the HER2-negative standard was IHC 0.

Other molecular marker tests

Hormone receptors, including estrogen receptor (ER) and progesterone receptor (PR), and Ki67 test standards used were based on the ASCO/CAP guidelines for ER and PR testing^[6] and were applied according to CAP guidelines^[7] for testing of biomarkers of breast cancer.

Antineoplastic protocols

Breast surgery includes breast-conserving therapy and mastectomy, and axillary surgery includes sentinel lymph node biopsy and axillary lymph node dissection of the level I and II lymph nodes of the ipsilateral axilla.^[8] Systematic treatment was conducted with reference to the National Comprehensive Cancer Network (NCCN) guidelines. In accordance with adjuvant/(neoadjuvant) chemotherapy regimens, as recommended by the guidelines, ≥ 4 treatment cycles were conducted for patients with breast cancer meeting the indications, and adjuvant endocrinotherapy for ≥ 5 years was conducted for patients with hormone receptor-positive breast cancer and patients

who were still receiving endocrinotherapy at the end of the follow-up period.

Follow-up

The main endpoint of the study was 5-year proportional disease-free survival (DFS). DFS was measured from the date when the patient received surgery to the date of the recurrence and metastasis or to the last follow-up. The secondary endpoint was 5-year proportional overall survival (OS). OS was measured from the date of diagnosis of breast cancer to death from any cause or to the last follow-up. All patients were followed up every 6 months, and the last follow-up date was May 2021. Follow-ups included breast and axillary lymph node B ultrasound, abdominal B ultrasound/computed tomography (CT), chest X-ray/chest CT, and other necessary examinations.

Statistical methods

SPSS (version 26.0; IBM, Armonk, NY, USA) was used for data processing. Measurement data are described as median deviation (min, max), and count and grade data are described as number of cases and percentages. The chi-squared test and Mann-Whitney U test were used to analyze the relationship between HER2 status and clinicopathological characteristics. Kaplan-Meier and Cox multiple regression analyses were used to analyze univariate and multivariate survival, 5-year DFS, and OS. The t test was used for the univariate analysis of clinicopathological features, and clinicopathological features with $P < 0.05$ were selected as valid features for inclusion in the machine learning model. All tests were two-sided tests unless otherwise stated, and $P < 0.05$ is considered statistically significant.

Recurrence prediction model

Data were discarded according to the inclusion criteria. Accordingly, cases with missing effective factors or cases with < 5 years of follow-up and without metastasis were removed. Cases with effective recurrence and metastasis were allocated to the training set and the test set at a ratio of 2:1, and the test set included effective cases with recurrence and metastasis and effective cases without recurrence and metastasis at a ratio of 1:1; the remaining effective cases without recurrence were assigned to the training set [Figure 1]. The Scikit-learn machine learning tool of Python software (<https://www.python.org/>) was used to construct recurrence prediction models. Six prediction models were constructed using the random forest model, support vector machine (SVM), k-nearest neighbor (KNN) method, logistic regression, naive Bayesian model (NBM), and AdaBoost. Grid search was used to determine the optimal hyperparameter of each model resulting in the best efficiency. The best prediction model was selected by comparing the area under the receiving operator characteristic (ROC) curve (AUC). The model was initially assigned, the training set samples were trained, and the test set samples were used to conduct external tests of the model^[9] [Figure 1]. Model performance was evaluated using ROC curves and the sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV).

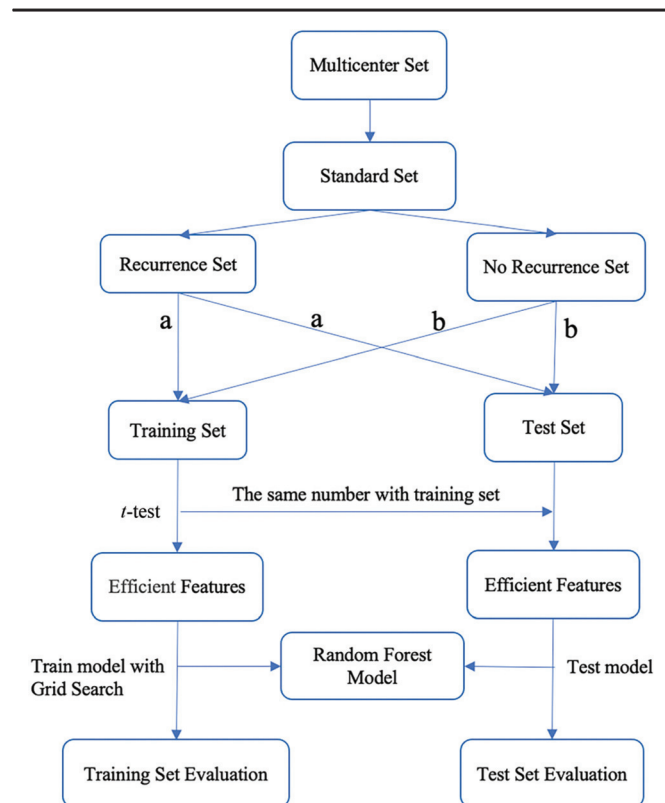


Figure 1: Flow chart of recurrence risk prediction model construction. (a) Cases with effective recurrence and metastasis were allocated to the training set and the test set at a ratio of 2:1. (b) The test set included effective cases with recurrence and metastasis and effective cases without recurrence and metastasis at a ratio of 1:1; the remaining effective cases without recurrence were assigned to the training set.

Results

Quality control of the HER2 test standard

In this study, clinical records were collected from 29 hospitals affiliated with the CSBrS, Chinese Society of Surgery of the Chinese Medical Association. Among them, 23 had pathology laboratories certified by PPCC and six were certified by ISO15189/CNAS-CL02 International Standardization Organization (ISO). All HER2 IHC test kits used by the 29 included hospitals were approved by NMPA: 23 hospitals used Ventana (Roche, USA), three used HercepTest™ (Dako, Denmark), 3 hospitals used domestic HER2 antibody provided by Genesd bio Co., Zsgb-bio Co and Anbipin Co. Ltd.^[10] All the participating hospitals used the NMPA HER2 fluorescence assay kits for HER2 FISH.

General information

From January 1, 2015 to December 31, 2016 a total of 25,096 women with early invasive breast cancer were treated at 29 academic hospitals. A total of 6547 cases of HER2-positive breast cancer were noted, accounting for 26.1% (17.50%–47.10%) of invasive breast cancers during that period; additionally, 7642 patients met the diagnostic standards for low-positive HER2 expression, accounting for 30.5% (7.80%–72.7%) of breast cancers

during that period. After ineligible cases were excluded, a total of 6486 patients met the inclusion criteria and were included in this study [Figures 2 and 3].

Clinical data analysis

A total of 6486 patients met the inclusion criteria for this study. Their median age was 50 years (20–90 years). The proportion of patients with low-positive HER2 expression ranged from 7.8% to 72.7% [Figure 2]. Among them, 3643 had HER2 IHC 1+ (56.2%), and 2843 had IHC 2+ (ISH negative) (43.8%). Significant differences were found between the two groups in tumor staging, histological type, histological grade, the Ki67 index, lymphovascular invasion, and hormone receptor status ($P < 0.05$); significant differences were also found between the two groups in the proportion of patients who received breast-conserving surgery, chemotherapy, and adjuvant endocrine therapy ($P < 0.01$) [Table 1].

Survival analysis

A total of 6486 patients with low-positive HER2 expression were followed up, with a median follow-up time of 57 months (4–76 months). A total of 506 (7.8%) cases of metastasis occurred during the follow-up, and the 5-year DFS was 92.1%; 167 (2.6%) patients died, and the 5-year OS was 97.4%. The univariate Cox regression analysis of recurrent and metastatic events found that age, tumor stage, lymphovascular invasion, the Ki67 index, histological grade, and hormone receptor status were related to prognosis ($P < 0.01$). Multivariate Cox regression analysis verified that tumor stage, lymphovascular invasion, and the Ki67 index were still related to prognosis ($P < 0.05$) [Table 2].

Recurrence risk prediction model

Machine learning methods were used to establish a recurrence risk prediction model for the 506 cases of low-positive HER2 expression cases in this study. The t test analysis obtained 15 effective features, including T staging, N staging, TNM staging, histological grade, vascular tumor thrombus, ER status, PR status, prognostic stage, Ki67 index (continuous variable), Ki67 index (grouping variable), chemotherapy, neoadjuvant chemotherapy, breast surgery, axillary surgery, and endocrine therapy ($P < 0.05$), which were included as independent variables in the constructed model [Table 3]. A total of 1915 patients with < 5 years of follow-up and 2835 patients with any effective feature missing were excluded. A total of 209 cases of recurrence and metastasis and 1534 cases without recurrence and metastasis were included in the study. Six prediction models were conducted using the random forest model, SVM, KNN, logistic regression, NBM, and AdaBoost [Figure 4]. The random forest model had the highest AUC and was chosen as the final recurrence prediction model. The final hyperparameters used to select the network were as follows: the number of trees in the forest was ten, the criterion used to measure the quality of a split was Gini, the random state was one, and the other parameters were set to default.

Discussion

Breast cancer is the most common malignant tumor in women. The literature reports 420,000 new breast cancer patients in China in 2020.^[11] Since the 21st century, with the continuous deepening of breast cancer awareness, breast cancer patients have been receiving specific treatment based on the classification of breast cancer. The molecular subtype of cancer not only is closely related to its prognosis but also serves as the basis for clinical treatment decisions. HER2-positive breast cancer has received extra attention in clinical practice due to its high invasiveness.^[12] In recent years, with the successful development of targeted drugs and the standardized use of anti-HER2 therapy, the prognosis of breast cancer patients with HER2 positive has markedly improved.^[13] In 2016, the 8th edition of the American Joint Committee on Cancer regarding breast cancer staging clearly stated that HER positivity is no longer a poor prognostic factor when targeted therapy is applied.^[14] On this basis, further in-depth research on issues related to HER2 expression levels is becoming a new clinical hotspot. In 2021, the NCCN Clinical Practice Guidelines for Breast Cancer and the CSCO Clinical Diagnosis and Treatment Guidelines voiced opinions regarding focusing on breast cancers with low-positive HER2 expression.^[15] To this end, the CSBrS initiated a multicenter study on early-stage breast cancer with low-positive HER2 expression to analyze clinicopathological information affecting its prognosis in China and to explore the establishment of a risk prediction model.

Standardization of HER2 diagnostic criteria is an important basis for making correct treatment decisions in clinical practice. This study required the participating hospitals to have laboratory qualifications such as the China National PQCC or International Organization for Standardization/China Conformity Assessment Country (ISO15189/CNAS-CL02) certification, to follow the 2013 ASCO/CAP requirements for breast cancer HER2 detection, and to use CFDA-approved IHC and fluorescence hybridization kits for HER2 detection. This study collected the data of 25,096 new early-stage breast cancer patients. Among them, HER2-positive breast cancer (6547 cases) accounted for 26.1% of invasive breast cancers during the study period, which is consistent with literature reports,^[16] indicating reliability of data from participating hospitals.

In recent years, the research and development of new ADCs has attracted considerable attention, and HER2 IHC 1+ and 2+ and ISH-negative cancers have been widely recognized as having low-positive HER2 expression. A phase II clinical trial showed that an ADC drug, DS-8201a, can achieve sustained lesion reduction in patients with advanced breast cancer and persistent low-positive HER2 expression despite multiple lines of rescue treatment^[3]; thus, it offers good prospects for expanding the indications for and potential benefits of HER2 treatment while simultaneously emphasizing the clinical need for greater refinement of HER2 status evaluations. The reported proportion of low-positive HER2 expression breast cancer was from 45% to 55%,^[17] and different

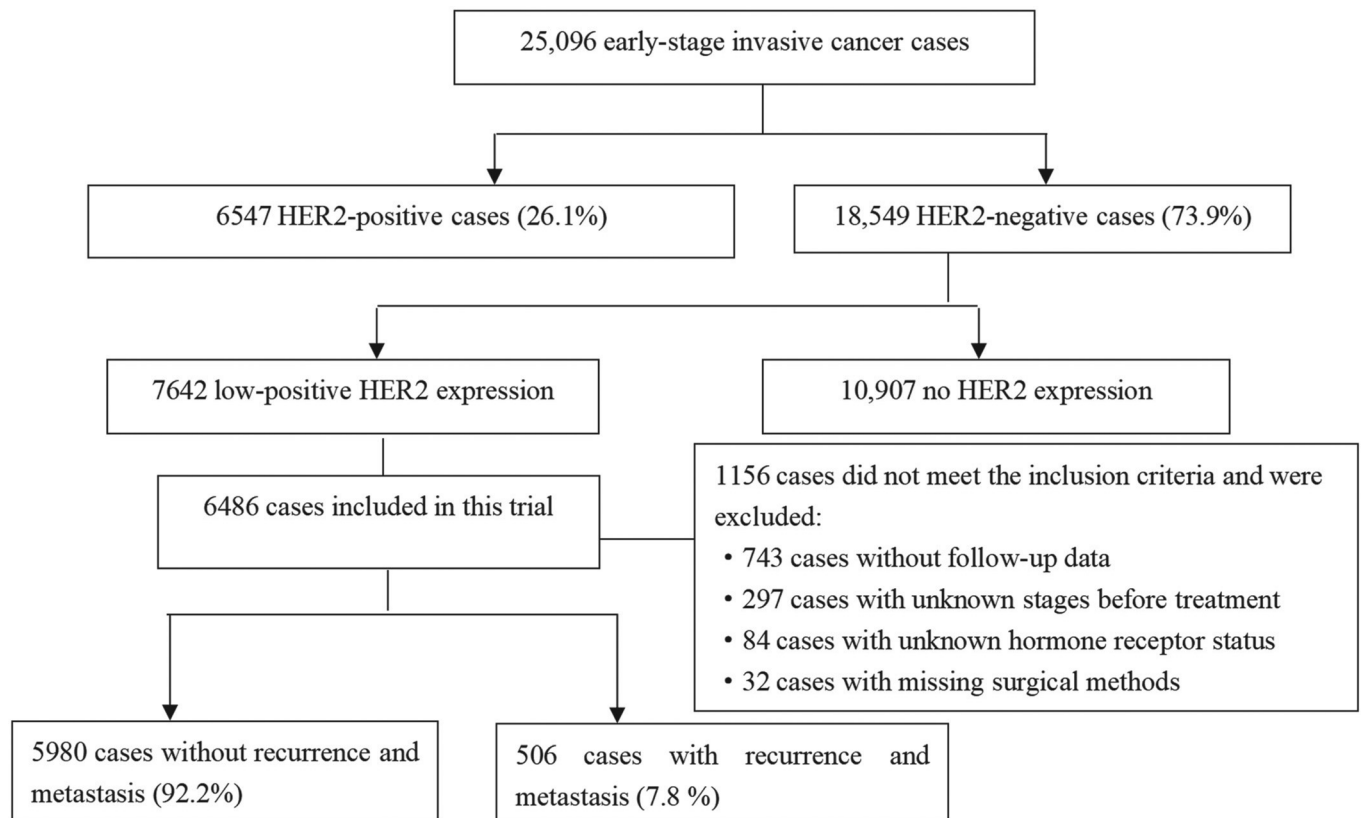


Figure 2: Inclusion process. HER2: Human epidermal growth factor receptor 2.

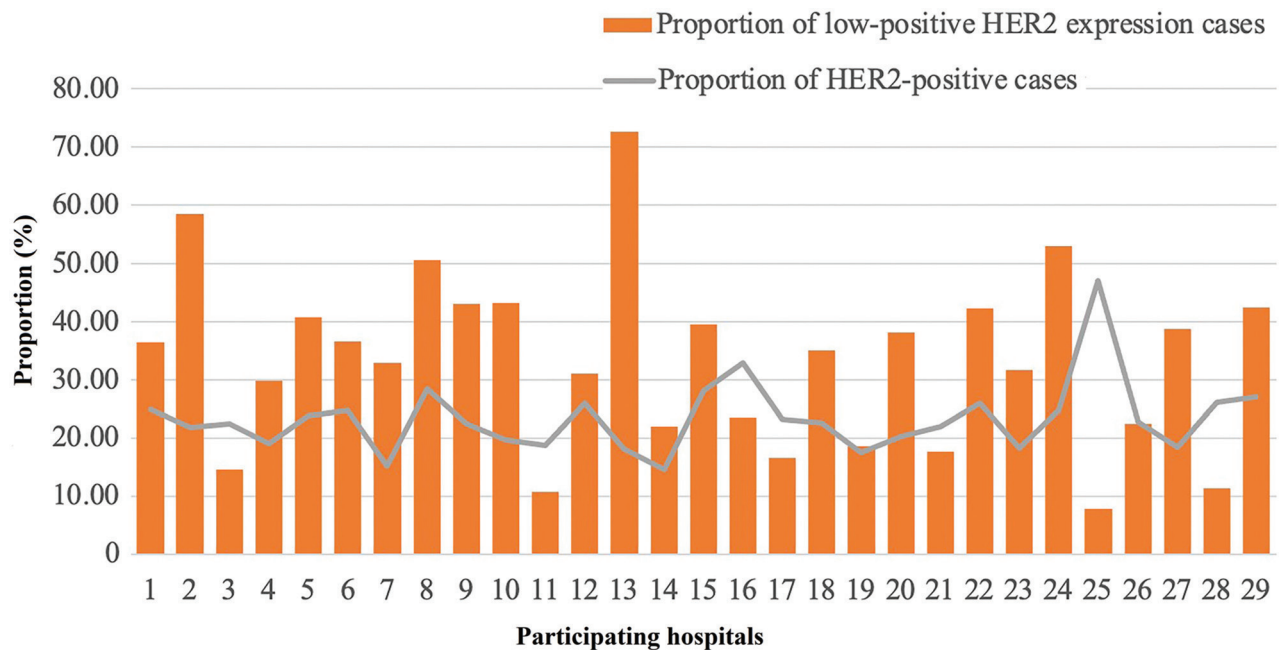


Figure 3: Proportion of low-positive HER2 expression and HER2-positive cases at participating hospitals. HER2: Human epidermal growth factor receptor 2.

detection kits used had an influence on the proportion of low-positive HER2 expression breast cancer. Hercep-Test™ (Dako, Denmark) classed several patients as IHC 0 that are HER2 IHC1 +/2 + (ISH-) by VENTANA (Roche, USA).^[18] In this study, the HER2 status of the included

patients was evaluated according to the 2013 ASCO/CAP standard. A total of 7642 patients with low-positive HER2 expression were included, accounting for 30.5% of the total breast cancer patients in the same period, which was lower than other reported proportions. A possible

Table 1: Clinicopathological data of HER2 IHC 1+ and IHC 2+ (ISH negative) breast cancer patients.

Characteristics	HER2 IHC 1 +	HER2 IHC 2+ (ISH negative)	χ^2	P value
Median age (years)	50 (20, 90)*	50 (22, 90)*	-0.470	0.638
Menstrual status			0.411	0.522
Menopausal	1702 (46.7)	1351 (47.5)		
Premenopausal	1941 (53.8)	1492 (52.5)		
T stage			-3.134 [†]	0.001
T0	2 (0.1)	0		
T1	1859 (51.0)	1324 (46.6)		
T2	1585 (43.5)	1382 (48.6)		
T3	160 (4.4)	107 (3.8)		
T4	37 (1.0)	30 (1.0)		
N stage			-4.020 [†]	<0.001
N0	2163 (59.4)	1531 (53.9)		
N1	1001 (27.4)	919 (32.3)		
N2	324 (8.9)	247 (8.7)		
N3	155 (4.3)	146 (5.1)		
TNM staging			-3.388 [†]	0.001
Stage I	1284 (35.2)	879 (30.9)		
Stage II	1814 (49.8)	1505 (52.9)		
Stage III	545 (15.0)	459 (16.2)		
Histological type			11.796	0.003
IDC-NOS	3258 (89.4)	2605 (91.6)		
ILC	136 (3.7)	68 (2.4)		
Other types	249 (6.9)	170 (6.0)		
Histological grade			-3.412 [†]	0.002
G1	289 (10.1)	161 (7.7)		
G2	2121 (74.0)	1551 (73.8)		
G3	457 (15.9)	389 (18.5)		
Missing	776	742		
Lymphovascular invasion			5.564	0.018
Positive	480 (18.8)	465 (21.6)		
Negative	2069 (81.2)	1688 (78.4)		
Missing	1094	690		
ER status			24.461	<0.001
Positive	3054 (83.8)	2507 (88.2)		
Negative	589 (16.2)	336 (11.8)		
PR status			16.947	<0.001
Positive	2756 (75.7)	2274 (80.0)		
Negative	887 (24.3)	569 (20.0)		
Prognostic staging			-1.895 [†]	0.058
Stage I	1576 (55.0)	1095 (52.1)		
Stage II	448 (15.6)	350 (16.7)		
Stage III	843 (29.4)	656 (31.2)		
Missing	776	742		
Ki67 index			-4.214 [†]	<0.001
<15%	929 (32.4)	549 (26.1)		
15%–30%	828 (28.8)	653 (31.1)		
>30%	1113 (38.8)	900 (42.8)		
Missing	773	741		
Chemotherapy			18.517	<0.001
Yes	2813 (77.6)	2317 (81.6)		
No	813 (22.4)	521 (18.4)		
Missing	17	5		
Neoadjuvant therapy			2.585	0.108
Yes	522 (18.6)	390 (16.8)		
No	2291 (81.4)	1927 (83.2)		
Missing or no chemotherapy	830	526		
Breast surgery			10.226	0.001
Breast-conserving surgery	905 (24.8)	610 (21.5)		
Total mastectomy	2738 (75.2)	2233 (78.5)		

Table 1
(continued).

Characteristics	HER2 IHC 1 +	HER2 IHC 2+ (ISH negative)	χ^2	P value
Axillary surgery			8.941	0.063
SLN negative, not dissected	1320 (36.2)	980 (34.5)		
Lymph node metastasis, ALND	458 (12.6)	419 (14.7)		
SLN positive, not ALND	43 (1.2)	42 (1.5)		
Direct dissection	1800 (49.4)	1380 (48.5)		
Not performed	22 (0.6)	22 (0.8)		
Radiotherapy			0.030	0.863
Yes	1692 (48.0)	1311 (47.8)		
No	1832 (52.0)	1432 (52.2)		
Missing	119	100		
Endocrine therapy			9.906	0.002
Yes	2947 (82.7)	2398 (85.6)		
No	618 (17.3)	404 (14.4)		
Missing	78	41		

*Median age (min, max). †Mann-Whitney U test. ALND: Axillary Lymph Node Dissection; ER: Estrogen receptor; HER2: Human epidermal growth factor receptor 2; IDC-NOS: Invasive breast carcinoma of no special type; IHC: Immunohistochemistry; ILC: Invasive lobular carcinoma; ISH: In situ hybridization; PR: Progesterone receptor; SLN: Sentinel lymph node; TNM: Tumour (T), regional lymph nodes (N), and metastatic involvement (M).

Table 2: Analysis of prognostic factors related to breast cancer recurrence and metastasis in cases of low-positive HER2 expression.

Variable	Univariate analysis		Multivariate analysis	
	Hazard ratio (95% CI)	P value	Hazard ratio (95% CI)	P value
HER2 status				
HER2 2+ (ISH-) vs. HER2 1+	0.925 (0.770–1.111)	0.403		
Age				
35 years vs. ≤35 years	0.595 (0.443–0.800)	0.001	0.747 (0.472–1.185)	0.747
Menstrual status				
Menopausal vs. premenopausal	1.007 (0.840–1.207)	0.938		
T stage				
T3–T4 vs. T0–T2	4.720 (3.731–5.970)	<0.001	3.142 (2.181–4.526)	<0.001
N stage				
N1–N3 vs. N0	2.838 (2.340–3.440)	<0.001	1.289 (0.894–1.860)	0.174
Histological grade				
G2–G3 vs. G1	1.575 (1.210–2.051)	0.001	1.005 (0.725–1.394)	0.974
Lymphovascular invasion				
Yes vs. No	2.418 (1.930–3.031)	<0.001	1.684 (1.260–2.251)	<0.001
Ki67 index				
>30% vs. ≤30%	1.668 (1.334–2.085)	<0.001	1.411 (1.067–1.867)	0.016
HR status				
HR negative vs. HR positive	1.622 (1.297–2.028)	<0.001	1.100 (0.765–1.581)	0.606
TNM staging				
Stage II–III vs. Stage I	4.095 (3.407–4.922)	<0.001	2.069 (1.438–2.978)	<0.001
Prognostic staging				
Stage II–III vs. Stage I	3.537 (2.739–4.568)	<0.001	1.478 (1.016–2.150)	0.041

CI: Confidence interval; HER2: Human epidermal growth factor receptor 2; HR: Hormone receptor; ISH: In situ hybridization.

reason for this discrepancy is that HER2 IHC test kits used by 21% (6/29) of participating hospitals were not VENTANA (Roche, USA). An additional reason may be that some patients with IHC 2+ breast cancer with low-positive HER2 expression were not included in the analysis because they did not undergo ISH testing during the study period. Besides, many patients with missing follow-up information were eliminated from the group,

resulting in lower proportion of low-positive HER2 expression breast cancer than reported.

The prognosis of early-stage breast cancer patients with low-positive HER2 expression is a matter of widespread concern. It is reported the local recurrence-free rates of patients with low-positive HER2 expression are lower than those of patients with HER2-zero breast cancer.^[19]

Table 3: Effective features of a 5-year recurrence risk model for breast cancer with low-positive HER2 expression.

Variable	t test	P value
T stage	-6.6856792	3.16×10^{-11}
N stage	-7.0331876	2.98×10^{-12}
TNM staging	-7.4456521	1.57×10^{-13}
Histological grade	-3.0961791	0.002
Lymphovascular invasion	5.8209306	7.05×10^{-9}
ER status	-2.2399913	0.025
PR status	-3.0788973	0.002
Prognostic stage	-5.9334610	3.63×10^{-9}
Ki67 index (continuous variable)	-4.7105960	2.68×10^{-6}
Ki67 index (grouping variable)	-3.9762508	7.31×10^{-5}
Chemotherapy	2.5179750	0.012
Neoadjuvant chemotherapy	6.2394494	5.61×10^{-10}
Breast surgery	-3.7477011	1.85×10^{-4}
Axillary surgery	-5.3146575	1.22×10^{-7}
Endocrine therapy	-2.2796274	0.022

ER: Estrogen receptor; HER2: Human epidermal growth factor receptor 2; PR: Progesterone receptor; TNM: Tumour (T), regional lymph nodes (N), and metastatic involvement (M).

Table 4: AUCs of recurrence risk prediction models using different methods.

Model	AUC of test set
Random forest model	0.8146
SVM	0.6420
KNN	0.7348
Logistic regression	0.7113
NBM	0.6928
AdaBoost	0.6023

AUC: Area under the receiving operator characteristic curve; CI: Confidence interval; KNN: K-nearest neighbor; NBM: Naive Bayesian model; SVM: Support vector machine.

In addition, the 3-year DFS and OS rates of patients with lowpositive HER2 expression have been found to be higher.^[20] In this study, HER2 status was evaluated with reference to the 2013 ASCO/CAP criteria. The follow-up data of 7642 early-stage breast cancer patients with low-positive HER2 expression revealed a 5-year DFS of 92.1% and a 5-year OS of 97.4%, suggesting that the overall prognosis of this type of early-stage breast cancer is good. At the same time, no significant difference in prognosis was found between the IHC 1 + and IHC 2 + /ISH-negative patients ($P = 0.403$), which is consistent with the conclusion of the subgroup analysis of the NSABP-B47 study.^[21] Significant differences in tumor staging, histological type, histological grade, the Ki67 index, lymphovascular invasion, and hormone receptor status were observed between the two groups ($P < 0.05$). Significant differences were also found in the proportion of patients who underwent breastconserving surgery and systemic treatments such as chemotherapy and endocrinotherapy ($P < 0.01$). The reasons need to be further studied. The univariate analysis of cases with metastasis found that age, tumor stage, lymphovascular invasion, the Ki67 index, histological grade, and hormone receptor status were correlated with prognosis ($P < 0.01$). The multivariate analysis verified that tumor stage, lymphovascular invasion, and the Ki67 index were still correlated with the prognosis ($P < 0.05$), suggesting that the tumor burden and invasiveness are important prognostic factors for breast cancer with lowpositive HER2 expression.

In recent years, machine learning methods have provided important help for establishing tumor clinical prognosis models.^[22,23] Machine learning is a multidisciplinary field involving probability theory, statistics, convex analysis, computer science, and other disciplines. The core idea of machine learning is based on various mathematical backgrounds using the increasing computing power of computers to analyze valuable mathematical laws from data and produce effective theories for practical work guidance. Machine learning can be categorized in many ways. In different application scenarios, selecting an appropriate model according to the data distribution is important to achieve the best prediction effect. At present, among the machine learning algorithms, some with excellent performance receive more attention. Compared with traditional regression algorithms, machine learning algorithms such as SVMs, decision trees, Bayesian net-

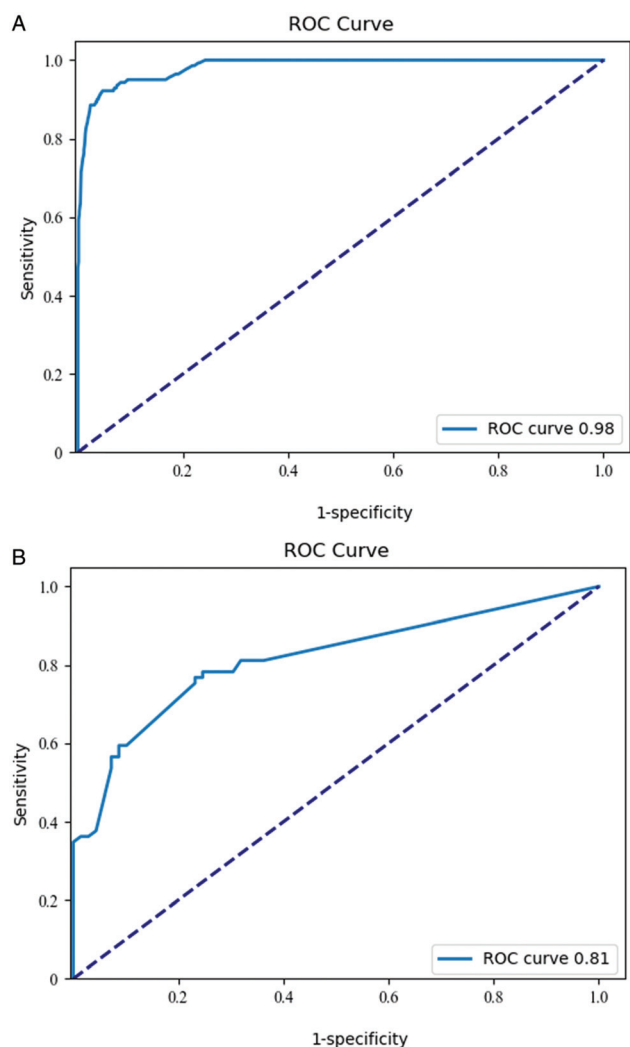


Figure 4: The ROC curves of the 5-year recurrence risk models for the training set and test set for low-positive HER2 expression breast cancer. (A) ROC curve of the recurrence risk model, training set. (B) ROC curve of the recurrence risk model, test set. HER2: Human epidermal growth factor receptor 2; ROC: Receiving operator characteristic.

Table 5: Machine learning model for predicting the 5-year recurrence risk in breast cancer patients with low-positive HER2 expression.

Data set	AUC	95% CI	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
Training set	0.983	0.977–0.990	98.6	77.9	29.9	99.8
Test set	0.815	0.750–0.880	78.3	71.0	73.0	76.6

AUC: Area under the receiving operator characteristic curve; CI: Confidence interval; HER2: Human epidermal growth factor receptor 2; NPV: Negative predictive value; PPV: Positive predictive value.

works, KNN algorithms, conditional random fields, and lifting methods have a stronger fit ability and can perform data distribution rule analysis. Because machine learning methods are based on different mathematical principles, they have different advantages. SVMs have advantages in various kinds of linearly indivisible data sets due to their multiple kernel functions. In scenarios with independent features, a naive Bayesian network shows higher performance for processing individual abnormal data, while for a random forest, the “randomness” in its design theory confers the ability to effectively balance errors in data sets with unbalanced data proportions, thus analyzing the importance of each feature, and it also has a high value for determining the correlation between the features.^[23] Because the data in this study are from patients from multiple centers, the amount of data is large, and the distribution of the data is uneven. Machine learning methods are used for data processing. After excluding cases with any feature missing, a total of 209 cases of recurrence and metastasis and 1534 cases without recurrence and metastasis were included in the study. T tests identified 15 effective features, including T stage, N stage, TNM staging, histological grade, lymphovascular invasion, ER status, PR status, prognostic stage, Ki67 index (continuous variable), Ki67 index (grouping variable), chemotherapy, neoadjuvant chemotherapy, breast surgery, axillary surgery, and endocrine therapy ($P < 0.05$), which were included as independent variables for model construction. Decision trees, random forests, SVMs, K nearest neighbors, and other machine learning algorithms were used to preprocess the data [Table 4]. The random forests were found to effectively eliminate irrelevant factors and analyze correlations between effective features and predicted values to identify differences and commonalities in data. The random forests fit best for the recurrence prediction model in our study.^[9,24] AUC of recurrence risk prediction model in training set was 0.983 (95% confidence interval [CI]: 0.977–0.990), and the test set AUC was 0.815 (95% CI: 0.750–0.880). The sensitivity of the 5-year recurrence risk prediction model for breast cancer with low-positive HER2 expression was 78.3%, and the specificity was 71.0%. The PPV of the model was 73.0%, and the NPV was 76.6% [Table 5], indicating that the model is valuable for identifying recurrence and metastasis in early-stage breast cancer with low-positive HER2 expression. This study provides a new reference for refining classified treatment for early-stage breast cancer with low-positive HER2 expression.

The definition of low-positive HER2 expression breast cancer has led to innovations in classified treatment. Machine learning is valuable for predicting the risk of

recurrence in breast cancer with low-positive HER2 expression and provides a reference for redefining the indications for anti-HER2 therapy.

Acknowledgements

Zhigang Yu, The Second Hospital of Shandong University; Hua Kang, Xuanwu Hospital, Capital Medical University; Ailing Song, Lanzhou University Second Hospital; Xingsong Tian, Shandong Provincial Hospital; Wei Zhu, Zhongshan Hospital, Fudan University; Hongchuan Jiang, Beijing Chaoyang Hospital, Capital Medical University; Kejin Wu, Obstetrics and Gynecology Hospital of Fudan University; Xiang Qu, Beijing Friendship Hospital affiliated to Capital Medical University; Erwei Song, Sun Yat-Sen Memorial Hospital; Yunjiang Liu, The Fourth Hospital of Hebei Medical University; Dahua Mao, Affiliated Wudang Hospital of Guizhou Medical University; Zhongwei Cao, Inner Mongolia Autonomous Region People’s Hospital; Jinghua Zhang, Tang Shan People’s Hospital; Yonghui Luo, The Second Affiliated Hospital of Nanchang University; Rui Ling, Xijing Hospital, Fourth Military Medical University; Zhenzhen Liu, Henan Tumor Hospital; Jian Huang, The Second Affiliated Hospital of Zhejiang University; Chuan Wang, Fujian Medical University Union Hospital; Shu Wang, Peking University People’s University; Jinping Liu, Sichuan Provincial People’s Hospital; Feng Jin, The First Hospital of China Medical University; Zhimin Fan, The First Hospital of Jilin University; Zuowei Zhao, The Second Affiliated Hospital of Dalian Medical University; Yi Zhao, Shengjing Hospital of China Medical University; Ke Liu, Jilin Cancer Hospital; Dedian Chen, Yunnan Cancer Hospital; Peifen Fu, The First Affiliated Hospital of Zhejiang University; Jun Jiang, The First Hospital Affiliated to Army Medical University/Southwest Hospital; Jiandong Wang, The First Medical Center of Chinese PLA General Hospital; Shui Wang, Jiangsu Province Hospital; Jianghua Ou, The First Affiliated Hospital of Xinjiang Medical University. Thanks for the help from Jun Yi Tai He Medical Technology Co. Ltd for establishing recurrence risk prediction model.

Funding

This study was supported by grants from the Youth Cultivation Fund of Beijing Medical Ward Foundation (No. 20180502) and Beijing Medical Ward Foundation (No. YXJL-2020-0941-0736).

Conflicts of interest

None.

References

1. Broglio KR, Quintana M, Foster M, Olinger M, McGlathlin A, Berry SM, *et al.* Association of pathologic complete response to neoadjuvant therapy in HER2-positive breast cancer with long-term outcomes: a meta-analysis. *JAMA Oncol* 2016;2:751–760. doi: 10.1001/jamaoncol.2015.6113.
2. Spring LM, Fell G, Arfe A, Sharma C, Greenup R, Reynolds KL, *et al.* Pathologic complete response after neoadjuvant chemotherapy and impact on breast cancer recurrence and survival: a comprehensive meta-analysis. *Clin Cancer Res* 2020;26:2838–2848. doi: 10.1158/1078-0432.Ccr-19-3492.
3. Modi S, Park H, Murthy RK, Iwata H, Tamura K, Tsurutani J, *et al.* Antitumor activity and safety of trastuzumab deruxtecan in patients with HER2-low-expressing advanced breast cancer: results from a phase Ib study. *J Clin Oncol* 2020;38:1887–1896. doi: 10.1200/jco.19.02318.
4. Wolff AC, Hammond MEH, Hicks DG, Dowsett M, McShane LM, Allison KH, *et al.* Recommendations for human epidermal growth factor receptor 2 testing in breast cancer: American Society of Clinical Oncology/College of American Pathologists Clinical Practice Guideline Update. *J Clin Oncol* 2013;31:3997–4013. doi: 10.1200/jco.2013.50.9984.
5. Breast Cancer guideline version 1.2015, NCCN Clinical Practice Guidelines in Oncology (NCCN Guidelines)[®]. Available from: <https://www.nccn.org/guidelines/guidelines-detail?category=1&cid=1419>. Accessed February 10, 2015]
6. Hammond MEH, Hayes DF, Wolff AC, Mangu PB, Temin S. American Society of Clinical Oncology/College of American Pathologists Guideline recommendations for immunohistochemical testing of estrogen and progesterone receptors in breast cancer. *J Oncol Pract* 2010;6:195–197. doi: 10.1200/jop.777003.
7. Fitzgibbons PL, Dillon DA, Alsabeh R, Berman MA, Hayes DF, Hicks DG, *et al.* Template for reporting results of biomarker testing of specimens from patients with carcinoma of the breast. *Arch Pathol Lab Med* 2014;138:595–601. doi: 10.5858/arpa.2013-0566-CP.
8. Jiao DC, Zhu JJ, Qin L, Guo XH, Zhao YJ, Chen XC, *et al.* Clinical practice guidelines for modified radical mastectomy of breast cancer: Chinese Society of Breast Surgery (CSBr) practice guidelines. *Chin Med J (Engl)* 2021;134:895–897. doi: 10.1097/cm9.0000000000001412.
9. Li H. *Statistical Learning*. 1st ed. Beijing: Tsinghua University Press; 2019.
10. Catalogue of Medical devices of NMPA. Available from: <https://www.nmpa.gov.cn/datasearch/home-index.html?itemId=2c9ba384759c957701759cd6a12803bc&79QlcAyHig6m=1642864769148#category=ylqx>. [Accessed July 21, 2021]
11. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, *et al.* Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2021;71:209–249. doi: 10.3322/caac.21660.
12. Seshadri R, Fergair FA, Horsfall DJ, McCaul K, Setlur V, Kitchen P. Clinical significance of HER-2/neu oncogene amplification in primary breast cancer. The South Australian Breast Cancer Study Group. *J Clin Oncol* 1993;11:1936–1942. doi: 10.1200/JCO.1993.11.10.1936.
13. Pondé N, Brandão M, El-Hachem G, Werbrouck E, Piccart M. Treatment of advanced HER2-positive breast cancer: 2018 and beyond. *Cancer Treat Rev* 2018;67:10–20. doi: 10.1016/j.ctrv.2018.04.016.
14. Amin MB, Edge S, Greene F, Byrd DR, Brookland RK, Washington MK, *et al.* *AJCC Cancer Staging Manual*. 8th ed. New York: Springer; 2016. 589–628.
15. Chinese Society of Clinical Oncology. *Breast Cancer Guidelines of Chinese Society of Clinical Oncology (CSCO)*. 1st ed. Beijing: People's Medical Publishing House; 2021. 16.
16. Waks AG, Winer EP. Breast cancer treatment: a review. *JAMA* 2019;321:288–300. doi: 10.1001/jama.2018.19323.
17. Schalper KA, Kumar S, Hui P, Rimm DL, Gershkovich P. A retrospective population-based comparison of HER2 immunohistochemistry and fluorescence in situ hybridization in breast carcinomas: impact of 2007 American Society of Clinical Oncology/College of American Pathologists criteria. *Arch Pathol Lab Med* 2014;138:213–219. doi: 10.5858/arpa.2012-0617-OA.
18. Scott M, Vandenberghe ME, Scorer P, Boothman AM, Barker C. Prevalence of HER2 low in breast cancer subtypes using the VENTANA anti-HER2/neu (4B5) assay. *J Clin Oncol* 2021;39:1021. doi: 10.1200/JCO.2021.39.15_suppl.1021.
19. Gilcrease MZ, Woodward WA, Nicolas MM, Corley LJ, Fuller GN, Esteva FJ, *et al.* Even low-level HER2 expression may be associated with worse outcome in node-positive breast cancer. *Am J Surg Pathol* 2009;33:759–767. doi: 10.1097/PAS.0-b013e31819437f9.
20. Denkert C, Seither F, Schneeweiss A, Link T, Blohmer JU, Just M, *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;22:1151–1161. doi: 10.1016/s1470-2045(21)0030101506.
21. Fehrenbacher L, Cecchini RS, Geyer CE Jr, Rastogi P, Costantino JP, Atkins JN, *et al.* NSAB B-47/NRG oncology phase III randomized trial comparing adjuvant chemotherapy with or without trastuzumab in high-risk invasive breast cancer negative for HER2 by FISH and with IHC 1+ or 2. *J Clin Oncol* 2020;38:444–453. doi: 10.1200/jco.19.01455.
22. Nie K, Shi L, Chen Q, Hu X, Jabbour SK, Yue N, *et al.* Rectal cancer: assessment of neoadjuvant chemoradiation outcome based on radiomics of multiparametric MRI. *Clin Cancer Res* 2016;22:5256–5264. doi: 10.1158/1078-0432.Ccr-15-2997.
23. Michelle TM. *Machine Learning*. 1st ed. New York: McGraw-Hill Education; 1997.
24. Harrington P. *Machine Learning in Action*. Beijing: Posts & Telecom Press; 2013.

How to cite this article: Xin L, Wu Q, Zhan C, Qin H, Xiang H, Xu L, Ye J, Duan X, Liu Y. Multicenter study of the clinicopathological features and recurrence risk prediction model of early-stage breast cancer with low-positive human epidermal growth factor receptor 2 expression in China (Chinese Society of Breast Surgery 021). *Chin Med J* 2022;135:697–706. doi: 10.1097/CM9.0000000000002056