Am J Cancer Res 2026;16(3):882-904
www.ajcr.us /ISSN:2156-6976/ajcr0171730

Original Article

Early recurrence predicts poor post-recurrence survival in young breast cancer patients: development and validation of a prognostic nomogram

Yun Cai[1], Yi Liu[1], Ye Sun[1], Yu Ren[2]

[1]Department of Traditional Chinese Medicine (TCM), The First Affiliated Hospital of Xi'an Jiaotong University, Xi'an 710061, Shaanxi, China; [2]Department of Breast Surgery, The First Affiliated Hospital of Xi'an Jiaotong University, Xi'an 710061, Shaanxi, China

Received January 28, 2026; Accepted March 18, 2026; Epub March 25, 2026; Published March 30, 2026

Abstract: This retrospective cohort study aimed to investigate clinicopathological differences between early and late recurrence in young breast cancer patients aged 40 years or younger, identify independent prognostic factors for post-recurrence overall survival (prOS), and develop and validate a nomogram model for predicting prOS. A total of 515 breast cancer patients initially diagnosed between January 2018 and January 2022 at The First Affiliated Hospital of Xi'an Jiaotong University were consecutively enrolled. All patients were aged 40 years or younger and experienced first recurrence during follow-up, which continued until January 2026. Patients diagnosed between January 2018 and December 2020 comprised the training cohort (n=294), while those diagnosed between January 2021 and January 2022 formed the temporal validation cohort (n=221). Early recurrence was defined as first recurrence within 24 months of initial treatment; late recurrence occurred beyond 24 months. The training and validation cohorts showed comparable baseline clinicopathological characteristics (all P>0.05). In the training cohort, patients with early recurrence exhibited more aggressive tumor features and higher rates of distant and visceral metastasis compared to those with late recurrence. Median prOS in the training cohort was 55 months, with significantly shorter prOS in the early recurrence group (P<0.001). Multivariate Cox regression identified the following as independent risk factors for prOS: early recurrence (hazard ratio [HR]=2.578, 95% confidence interval [CI]: 1.711-3.883), tumor size T3-T4 (HR=1.950, 95% CI: 1.321-2.878), lymph node positivity (HR=2.190, 95% CI: 1.512-3.172), human epidermal growth factor receptor 2 (HER2)-overexpressing subtype (HR=2.347, 95% CI: 1.255-4.387), triple-negative breast cancer (TNBC) (HR=2.293, 95% CI: 1.250-4.205), receipt of neoadjuvant chemotherapy (HR=1.782, 95% CI: 1.277-2.488), and age younger than 35 years (HR=1.424, 95% CI: 1.012-2.003). The nomogram was constructed based on these factors. In the training cohort, the area under the curve (AUC) for predicting 1-year, 3-year, and 5-year prOS was 0.845, 0.820, and 0.649, respectively, with a concordance index (C-index) of 0.756. In the validation cohort, AUC values were 0.720, 0.725, and 0.601, respectively, with a C-index of 0.658. Calibration curves and decision curve analysis demonstrated acceptable predictive accuracy and clinical utility. In conclusion, early recurrence serves as an important predictor of poor post-recurrence survival in young breast cancer patients aged 40 years or younger. The nomogram incorporating recurrence timing, tumor burden, and molecular subtype can predict prOS with reasonable accuracy and may guide clinical risk stratification and individualized management.

Keywords: Breast cancer, young patients, early recurrence, post-recurrence overall survival, nomogram, prognostic prediction

Introduction

Breast cancer is the most frequent cancer seen in women around the world and is a leading cause of cancer death in women. In 2020, there were approximately 2.3 million new cases of breast cancer and 685,000 deaths from this disease globally [1]. China is increasingly afflicted by this condition. In 2022, the number of new cases and death due to breast cancer in women of China were about 357000 and 75000, accounting for 15.59% and 7.94% of all female malignancies, respectively [2]. Notably, Chinese women suffering from breast cancer are diagnosed in their 45 to 54 years age group, a decade earlier than in the West [3]. The inci-

https://doi.org/10.62347/ZWIX3394

dence of breast cancers is increasing in young Chinese women (under the age of 40 years) according to recent epidemiological data. The 20-34 age-group in rural areas is witnessing an annual growth of up to 9.0% [4]. The rising incidence and changing epidemiology of breast cancer among young women necessitate research focused on this group, especially in China.

Breast cancer patients who are aged 40 and younger tend to account for 5%-7% of all breast cancer cases [5]. Young women with breast cancer have different biological behavior and clinicopathological characteristics than older patients. According to studies, younger patients are more often diagnosed with high-grade tumors that are hormone receptor negative, HER2 overexpressed, and triple-negative breast cancer (TNBC) [6]. An analysis of the United States Surveillance, Epidemiology, and End Results (SEER) database on a large scale showed that women aged below 40 were more likely to be diagnosed with the HER2-positive or triple-negative subtypes and at a more advanced stage (all $P<0.001$) [7]. Younger breast cancer patients have a worse prognosis than older patients, despite treatment advances. The populations have achieved only minor improvements in survival over recent decades [8]. The combined effects of these adverse biological and clinical features result in an elevated risk of recurrence and a worse prognosis over the long term in younger patients, necessitating an understanding of the patterns of recurrence and post-recurrence prognosis in this subset of patients for clinical relevance.

The recurrence of breast cancer has a serious impact on long-term survival. Around 25-30% of patients will have their disease recurring, and die of metastatic disease [9]. The timing of recurrent events is an important prognostic marker that is clinically defined as early versus late recurrence. The patterns of recurrence are different for different molecular subtypes. Triple-negative breast cancer (TNBC) are associated with an earlier recurrence, whereas hormone receptor-positive breast cancer are linked to ongoing late recurrence that may occur more than 20 years after diagnosis [10, 11]. A study of a large cohort of Danish breast cancer patients found that, amongst patients who lived a disease-free interval of 10 years, 16.6% experienced a recurrence between years 10 and 32. For those diagnosed before the age of 40 years, this was an independent risk factor for late recurrence [12]. Nonetheless, though these analyses have described the timing and risk of a recurrence, those did mainly predict an initial recurrence event. Recurrence in younger patients has not been studied as much. What happens after it? Once a recurrence has happened, the clinical course would be very different and the factors that would influence later survival would be quite different from those determining the first recurrence. The inability to accurately predict the clinical course of breast cancer in young adults can potentially result in treatment under- or overtreatment.

Post-recurrence overall survival (prOS) is an important prognostic factor in recurrent breast cancer. The timing of recurrence, the molecular subtype, the site of metastases, and response to treatment affect prOS [13]. Nonetheless, the majority of previous studies mainly predicted the recurrence risk in breast cancer broadly rather than in young patients, thus failing to delineate the prognostic significance of early or late recurrence. It remains unclear whether recurrence that occurs earlier may bear worse prognostic implications than later recurrence in this age group and whether the prognostic impact varies by molecular subtypes and metastatic sites. In addition, clinicopathological differences in early and late recurrence, prOS variation across molecular subtypes and recurrence sites, as well as precise prediction models for prOS in young patients require further exploration. No existing study has reported any nomogram specifically for predicting prOS in young patients with recurrent breast cancer. Nevertheless, nomograms are intuitive and practical tools that have been widely used for individualized risk prediction for several malignancies [14]. Due to specific biological behaviour and worse prognosis in young patients with breast cancer, a specific tool that predicts recurrence timing in conjunction with standard clinicopathological factors would add clinically meaningful value in individualized post-recurrence management.

With this background, our study aimed to compare clinicopathological characteristics between early and late recurrence in young breast cancer patients aged 40 years or younger, to

determine independent prognostic factors affecting prOS, and to develop and validate a nomogram predicting prOS. As far as we know, this study is the first to describe post-recurrence survival outcomes in young breast cancer patients by adding recurrence timing, tumor burden, and molecular subtype to a validated prognostication model. Furthermore, we aimed to use a validation cohort for making sure our results are generalizable to be beyond a single institution. We expect that the discovery of research-based evidence for tailored management and prognosis prediction of recurrent breast cancer in young patients will enhance risk stratification and clinical decision-making of patients in this scenario.

Methods and data

Study design and participants

The research used a retrospective cohort design. We enrolled breast cancer patients with initial diagnoses from January 2018 until January 2022 successively at The First Affiliated Hospital of Xi'an Jiaotong University. Patients aged 40 years or younger with pathologically confirmed breast cancer who experienced first recurrence during follow-up were eligible. Follow-up continued till January 2026 or death, whichever occurred first. A grand total of 515 patients were included. The training cohort (n=294) consisted of patients diagnosed from January 2018 to December 2020, while January 2021 to January 2022 patients made up the temporal validation cohort (n=221). This study was approved by the Institutional Ethics Committee of The First Affiliated Hospital of Xi'an Jiaotong University (Approval No.: XJTU1AF-2023LSL-020, December 2023). Informed consent was conducted according to ethics committee requirements given the study's retrospective nature (Figure S1).

Sample size and statistical power

The main objective of this study was overall survival in the patient with prostate cancer and the main outcome size was reported as hrr Sample size calculation was done using Schoenfeld event number methodology. The required number of death events (D) can be calculated according to the following formula: $D = (Z_{1-\alpha/2} + Z_{1-\beta})^2/([\ln(HR)]^2 \times p(1-p))$, where α is the two-sided significance level, 1-β is the statistical power, and p is the proportion of patients with early recurrence in the total population. If the proportion of patients destined to die from these causes is q, total sample size is derived as N = D/q. Previous research indicates that early recurrence is significantly associated with poor prognosis. According to relevant research and meta-analyses, effect sizes have been reported to usually fall within the HR range of 1.5-2.0 [11, 15, 16]. Younger patients also show differences in late recurrence risk [12]. In light of this evidence, we set HR=1.68 (based on [16]), for sample size estimation with α= 0.05, power 1-β=0.80, early recurrence proportion P=0.50, proportion of event death q=0.50 (based on [11]). The value of D≈117 events and N≈234 patients was obtained by substituting values ($Z_{1-\alpha/2}$=1.96, $Z_{1-\beta}$=0.84, ln(1.68)=0.519). Sensitivity analysis indicated that when HR=2.0 approximately 130 patients would be required. N ≈ 314 patients needed when power rises to 90% ($Z_{1-\beta}$=1.28) with HR=1.68. The final sample of 515 recurrent breast cancer patient exceeds the necessary maximum requirement across these scenarios (about 314) so we are powered enough to detect prOS differences resulting from early or late recurrence.

Inclusion and exclusion criteria

Inclusion criteria were: (1) Female patients aged 40 years or younger at initial diagnosis; (2) Pathologically confirmed invasive breast cancer by surgical excision or core needle biopsy; (3) Received standardized initial treatment (surgery with or without adjuvant/neoadjuvant therapy) at the study institutions; (4) Experienced first documented recurrence (locoregional recurrence or distant metastasis) during follow-up, confirmed by imaging, pathology, or clinical evidence; (5) Complete baseline clinicopathological data and follow-up records available for statistical analysis; (6) Follow-up duration from initial diagnosis to first recurrence or last contact was at least 3 months.

Exclusion criteria were: (1) Stage IV (de novo metastatic) disease at initial diagnosis; (2) Ductal carcinoma in situ (DCIS) or other non-invasive breast pathology without invasive component; (3) History of other primary malignancies (excluding cured non-melanoma skin cancer or cervical carcinoma in situ); (4) Bilateral breast cancer at initial diagnosis; (5) Male

breast cancer; (6) Recurrence status unconfirmed or unable to determine the timing of first recurrence; (7) Unable to ascertain survival outcome (lost to follow-up immediately after recurrence with no survival information available); (8) Missing key variables (including molecular subtype, tumor size, or lymph node status) that could not be supplemented through medical record review.

Data collection and variable definitions

Demographic characteristics, tumor features, treatment information, and recurrence/metastasis patterns were extracted from electronic medical records and follow-up data. Demographic and recurrence-related variables included age (stratified as ≤35 years vs. 36-40 years), recurrence type (early vs. late recurrence, with early recurrence defined as first recurrence within 24 months of initial treatment and late recurrence as beyond 24 months), and prOS (time from first recurrence confirmation to death or last follow-up).

Tumor clinicopathological characteristics included tumor size (T1-T2, T3-T4), histological grade (I, II, III), lymph node status (N0, N1-N3), estrogen receptor (ER) status, progesterone receptor (PR) status, HER2 status, Ki-67 expression (low expression <30%, high expression ≥30%), and molecular subtype (Luminal A, Luminal B, HER2-overexpressing, TNBC). Treatment-related variables included surgical approach (mastectomy/breast-conserving surgery), neoadjuvant chemotherapy (yes/no), adjuvant chemotherapy (yes/no), adjuvant radiotherapy (yes/no), and adjuvant endocrine therapy (yes/no).

Recurrence/metastasis pattern variables were recorded at first recurrence and included distant metastasis, locoregional recurrence, bone metastasis, liver metastasis, lung metastasis, brain metastasis, visceral metastasis, and multi-organ metastasis (all yes/no). First recurrence sites were stratified by priority order: brain metastasis, visceral metastasis (non-brain), bone-only metastasis, and locoregional recurrence.

Measurement methods

Tumor clinicopathological characteristics, including tumor size, histological grade, and lymph node status, were determined based on pathological reports from surgical excision or core needle biopsy specimens obtained at the time of initial diagnosis. ER, PR, HER2, and Ki-67 status were assessed using immunohistochemistry on the same initial diagnostic specimens. Testing was performed by the pathology department. Immunohistochemical staining was conducted on Ventana BenchMark XT/ULTRA platforms (Roche Diagnostics, Basel, Switzerland). Antibody reagents included ER (clone SP1), PR (clone 1E2), HER2 (clone 4B5), and Ki-67 (clone MIB-1), all supplied by Roche/Ventana.

Treatment-related variables, including surgical approach, neoadjuvant chemotherapy, adjuvant chemotherapy, adjuvant radiotherapy, and adjuvant endocrine therapy, were recorded based on the treatment administered during the initial treatment period prior to disease recurrence.

First recurrence and metastatic sites were documented at the time of first recurrence confirmation through imaging examinations, pathological evidence, and comprehensive clinical assessment based on medical records and follow-up data. Recurrence type (locoregional recurrence or distant metastasis) and specific metastatic sites were recorded according to the primary manifestation at first recurrence.

Outcome measures

The primary outcome was prOS. Secondary outcomes included: comparability of baseline characteristics between training and validation cohorts; clinicopathological differences between early and late recurrence in the training cohort; prOS differences stratified by molecular subtype and first recurrence site; internal validation performance of the prediction model in the training cohort (discrimination, calibration, and clinical net benefit); external validation performance in the validation cohort; and prOS differences between risk groups based on risk score stratification.

Statistical analysis

All statistical analyses were performed using R 4.5.2 (R Foundation for Statistical Computing, Vienna, Austria) and SPSS 27.0 (IBM Corp., Armonk, NY, USA). Categorical variables were

expressed as counts (percentages) and compared using chi-square test or Fisher's exact test. Continuous variables were described as mean ± standard deviation or median (interquartile range) according to the data distribution, and compared using Student's t-test or Mann-Whitney U test as appropriate. The analysis of survival data utilized the Kaplan-Meier method and log-rank test for relevant comparisons of groups. Univariate and Multivariate analyses were performed using Cox proportional hazards regression and results were reported as hazard ratio (HR) and 95% confidence interval (CI). Using a Cox regression model for multivariable analysis, the variables selection strategy was as follows. First, univariate Cox regression analysis was conducted for all clinicopathological and treatment-related variables. Any variable that shows P<0.05 in univariate analysis will be considered a candidate for the multivariate model. Before entering the multivariate model, variance inflation factor (VIF) was calculated to assess and control multicollinearity between those variables. The absence of significant collinearity was indicated with VIF<5. Again, moderate collinearity was indicated with 5≤VIF<10. Finally, VIF≥10 indicated severe collinearity. Sequentially, variables with high collinearity were removed based on clinical relevance and VIF values until all VIFs of remaining variables were acceptable. The other candidate variables were entered using the backward stepwise selection method (likelihood ratio test) into the multivariate Cox proportional hazards regression model with an entry criterion of P<0.05 and removal criterion of P>0.10. Schoenfeld residuals and residual trend plots over a timeline were used to test the proportional hazards assumption. A global test P>0.05 suggested that overall the model satisfied the proportional hazards assumption. A nomogram was generated relying on the final multivariate Cox model. The concordance index (C-index) was calculated using Bootstrap resampling (n=1000) for internal validation, The discrimination of each model was assessed through time-dependent receiver operating characteristic (ROC) curves and area under the curve (AUC). The calibration curves, Brier score, and Hosmer-Lemeshow (H-L) test evaluated calibration. Decision curve analysis (DCA) was used in assessing clinical utility. The Cox model served as the basis for calculating individual risk scores. The median risk score (1.668) of the training cohort was used to classify both cohorts into high-risk and low-risk training and validation cohorts. The prognostic outcome group was compared using Kaplan-Meier curves. All tests were performed two-sided and were statistically significant at P<0.05.

Results

Comparability of baseline characteristics between training and validation cohorts

This study enrolled 515 recurrent breast cancer patients aged 40 years or younger, with 294 in the training cohort and 221 in the validation cohort. The two cohorts showed no statistically significant differences in recurrence type (P=0.918), age (P=0.373), tumor size (P=0.677), histological grade (P=0.905), lymph node status (P=0.068), ER status (P=0.283), PR status (P=0.714), HER2 status (P=0.964), Ki-67 expression (P=0.684), molecular subtype (P=0.730), surgical approach (P=0.679), neoadjuvant chemotherapy (P=0.458), adjuvant chemotherapy (P=0.847), adjuvant radiotherapy (P=0.880), adjuvant endocrine therapy (P=0.644), distant metastasis (P=0.502), locoregional recurrence (P=0.115), bone metastasis (P=0.589), liver metastasis (P=0.346), lung metastasis (P=0.966), brain metastasis (P=0.595), visceral metastasis (P=1.000), or multi-organ metastasis (P=0.340). All *P* values exceeded 0.05, demonstrating good comparability between cohorts (see Table S1).

Clinicopathological differences between early and late recurrence in the training cohort

Regarding demographic and tumor characteristics, the early recurrence group had a higher proportion of patients aged 35 years or younger (58.16% vs. 45.75%, P=0.033), more T3-T4 tumors (23.40% vs. 12.42%, P=0.014), more histological grade III tumors (51.06% vs. 31.37%, P<0.001), and higher lymph node positivity rates (77.30% vs. 60.13%, P=0.002).

Molecular markers were evaluated, the early recurrence group showed a significantly higher rate of ER negativity at 72.34% versus 39.87% in the late recurrence group (P<0.001). PR negativity was also higher at 76.60% versus 54.25% (P<0.001), as was the high Ki-67 expression at 53.19 versus 37.91% in the late

Table 1. Comparison of clinicopathological characteristics between early and late recurrence groups in the training cohort

Characteristic	Total	Early Recurrence (n=141)	Late Recurrence (n=153)	Statistic	*P* value
Age (years)				4.521	0.033
36-40 years	142 (48.30%)	59 (41.84%)	83 (54.25%)		
≤35 years	152 (51.70%)	82 (58.16%)	70 (45.75%)		
Tumor size				6.083	0.014
T1-T2	242 (82.31%)	108 (76.60%)	134 (87.58%)		
T3-T4	52 (17.69%)	33 (23.40%)	19 (12.42%)		
Histological grade				14.853	<0.001
Grade I	40 (13.61%)	11 (7.80%)	29 (18.95%)		
Grade II	134 (45.58%)	58 (41.13%)	76 (49.67%)		
Grade III	120 (40.82%)	72 (51.06%)	48 (31.37%)		
Lymph node status				10.008	0.002
N1-3	201 (68.37%)	109 (77.30%)	92 (60.13%)		
N0	93 (31.63%)	32 (22.70%)	61 (39.87%)		
ER status				31.318	<0.001
Negative	163 (55.44%)	102 (72.34%)	61 (39.87%)		
Positive	131 (44.56%)	39 (27.66%)	92 (60.13%)		
PR status				16.101	<0.001
Negative	191 (64.97%)	108 (76.60%)	83 (54.25%)		
Positive	103 (35.03%)	33 (23.40%)	70 (45.75%)		
HER2 status				3.151	0.076
Negative	187 (63.61%)	97 (68.79%)	90 (58.82%)		
Positive	107 (36.39%)	44 (31.21%)	63 (41.18%)		
Ki-67 expression				6.918	0.009
Low (<30%)	161 (54.76%)	66 (46.81%)	95 (62.09%)		
High (≥30%)	133 (45.24%)	75 (53.19%)	58 (37.91%)		
Molecular subtype				45.867	<0.001
Luminal A	57 (19.39%)	18 (12.77%)	39 (25.49%)		
Luminal B	94 (31.97%)	33 (23.40%)	61 (39.87%)		
HER2-overexpressing	66 (22.45%)	28 (19.86%)	38 (24.84%)		
TNBC	77 (26.19%)	62 (43.97%)	15 (9.80%)		
Surgical approach				0.082	0.774
Mastectomy	181 (61.56%)	88 (62.41%)	93 (60.78%)		
Breast-conserving surgery	113 (38.44%)	53 (37.59%)	60 (39.22%)		
Neoadjuvant chemotherapy				4.112	0.043
No	166 (56.46%)	71 (50.35%)	95 (62.09%)		
Yes	128 (43.54%)	70 (49.65%)	58 (37.91%)		
Adjuvant chemotherapy				0.075	0.784
No	20 (6.80%)	9 (6.38%)	11 (7.19%)		
Yes	274 (93.20%)	132 (93.62%)	142 (92.81%)		
Adjuvant radiotherapy				2.497	0.114
No	135 (45.92%)	58 (41.13%)	77 (50.33%)		
Yes	159 (54.08%)	83 (58.87%)	76 (49.67%)		
Adjuvant endocrine therapy				17.531	<0.001
No	163 (55.44%)	96 (68.09%)	67 (43.79%)		
Yes	131 (44.56%)	45 (31.91%)	86 (56.21%)		

Distant metastasis				3.940	0.047
No	67 (22.79%)	25 (17.73%)	42 (27.45%)		
Yes	227 (77.21%)	116 (82.27%)	111 (72.55%)		
Locoregional recurrence				14.889	<0.001
No	166 (56.46%)	96 (68.09%)	70 (45.75%)		
Yes	128 (43.54%)	45 (31.91%)	83 (54.25%)		
Bone metastasis				1.738	0.187
No	174 (59.18%)	89 (63.12%)	85 (55.56%)		
Yes	120 (40.82%)	52 (36.88%)	68 (44.44%)		
Liver metastasis				9.856	0.002
No	207 (70.41%)	87 (61.70%)	120 (78.43%)		
Yes	87 (29.59%)	54 (38.30%)	33 (21.57%)		
Lung metastasis				7.844	0.005
No	216 (73.47%)	93 (65.96%)	123 (80.39%)		
Yes	78 (26.53%)	48 (34.04%)	30 (19.61%)		
Brain metastasis				15.611	<0.001
No	265 (90.14%)	117 (82.98%)	148 (96.73%)		
Yes	29 (9.86%)	24 (17.02%)	5 (3.27%)		
Visceral metastasis				14.770	<0.001
No	149 (50.68%)	55 (39.01%)	94 (61.44%)		
Yes	145 (49.32%)	86 (60.99%)	59 (38.56%)		
Multi-organ metastasis				17.075	<0.001
No	261 (88.78%)	114 (80.85%)	147 (96.08%)		
Yes	33 (11.22%)	27 (19.15%)	6 (3.92%)		

Note: ER, Estrogen receptor; PR, Progesterone receptor; HER2, Human epidermal growth factor receptor 2; TNBC, Triple-negative breast cancer.

recurrence group (P=0.009). Analysis of molecular subtype showed that early recurrence (43.97%) prevalence of triple-negative breast cancer (TNBC) in contrast to late recurrence (25.49% Luminal A + 39.87% Luminal B) P< 0.001. In terms of management, more patients in the early recurrence group underwent the neoadjuvant chemotherapy (49.65% vs. 37.91%; P=0.043) whereas less received the adjuvant endocrine therapy (31.91% vs. 56.21%; P<0.001). The HER2 status (P=0.076), surgical approach (P=0.774), adjuvant chemotherapy (P=0.784), and adjuvant radiotherapy (P=0.114) were not significantly different between the two groups.

For recurrence and metastasis patterns, the early recurrence group had significantly higher rates of distant metastasis (82.27% vs. 72.55%, P=0.047), liver metastasis (38.30% vs. 21.57%, P=0.002), lung metastasis (34.04% vs. 19.61%, P=0.005), brain metastasis (17.02% vs. 3.27%, P<0.001), visceral metastasis (60.99% vs. 38.56%, P<0.001), and multi-organ metastasis (19.15% vs. 3.92%, P<0.001). The early recurrence group had lower locoregional recurrence rates (31.91% vs. 54.25%, P<0.001). Bone metastasis rates did not differ significantly between groups (36.88% vs. 44.44%, P=0.187) (see **Table 1**).

Survival analysis of post-recurrence overall survival in the training cohort

Median prOS in the 294 training cohort patients was 55 months. The 1-year, 2-year, 3-year, and 5-year prOS rates were 83.0%, 72.1%, 61.9%, and 46.3%, respectively.

Stratified analysis by recurrence timing revealed significant prOS differences between early and late recurrence groups (P<0.001). Median prOS was 32 months in the early recurrence group, while median prOS was not reached in the late recurrence group. Using the late recurrence group as reference, the early recurrence group showed significantly elevated mortality risk (HR=2.80, 95% CI: 1.99-3.95).

Analysis stratified by molecular subtype demonstrated significant prOS differences across subtypes (P<0.001). Compared to Luminal A, both TNBC (HR=3.83, 95% CI: 2.24-6.57) and HER2-overexpressing subtype (HR=2.78, 95% CI: 1.59-4.87) showed significantly increased mortality risk. Luminal B did not differ significantly from Luminal A (HR=1.28, 95% CI: 0.72-2.26). TNBC had the shortest median prOS at 27 months, followed by HER2-overexpressing subtype at 35.5 months. Neither Luminal A nor Luminal B reached median prOS.

Analysis stratified by first recurrence site also showed significant prOS differences (P=0.013). Using locoregional recurrence as reference, visceral metastasis (non-brain) patients had significantly higher mortality risk (HR=3.03, 95% CI: 1.46-6.29), as did bone-only metastasis patients (HR=2.26, 95% CI: 1.04-4.93). Brain metastasis patients did not differ significantly from locoregional recurrence patients (HR=2.14, 95% CI: 0.89-5.17). Median prOS for visceral metastasis (non-brain) was 46 months; median prOS was not reached in other groups (see **Figure 1**).

Proportional hazards assumption testing and multicollinearity diagnosis for cox regression

Variable coding is shown in **Table 2**. Initial VIF analysis of 23 variables revealed severe multicollinearity (VIF>10) for visceral metastasis (VIF=14.856), liver metastasis (VIF=13.502), lung metastasis (VIF=13.228), and molecular subtype (VIF=10.736). Moderate collinearity (5≤VIF<10) was observed for distant metastasis (VIF=7.541) and multi-organ metastasis (VIF=5.317).

After sequential removal of high-collinearity variables, 11 variables were retained for Cox regression: recurrence type, age, tumor size, histological grade, lymph node status, molecular subtype, surgical approach, neoadjuvant chemotherapy, adjuvant chemotherapy, adjuvant radiotherapy, and visceral metastasis. All adjusted VIF values were below 2 (range 1.049-1.386), indicating effective control of multicollinearity (see **Table 2**).

Schoenfeld residual analysis tested the proportional hazards assumption for these 11 variables. The global test indicated that the overall model satisfied the proportional hazards assumption (χ^2=21.173, df=14, P=0.097). Individual variable tests confirmed that recurrence type (P=0.486), age (P=0.192), tumor size (P=0.142), histological grade (P=0.155), lymph node status (P=0.073), molecular subtype (P=0.055), surgical approach (P=0.672), neoadjuvant chemotherapy (P=0.610), adjuvant chemotherapy (P=0.191), adjuvant radiotherapy (P=0.317), and visceral metastasis (P=0.649) all met the assumption (all P>0.05). Schoenfeld residual trend plots showed relatively stable residual patterns over time, further supporting the proportional hazards assumption (see **Figure 2**).

Cox regression analysis of post-recurrence overall survival in the training cohort

Univariate Cox regression identified several factors significantly associated with worse prOS: age younger than 35 years (HR=1.574, 95% CI: 1.128-2.196, P=0.008), early recurrence (HR=2.801, 95% CI: 1.985-3.953, P<0.001), tumor size T3-T4 (HR=2.640, 95% CI: 1.824-3.822, P<0.001), lymph node positivity (HR=1.581, 95% CI: 1.128-2.216, P=0.008), HER2-overexpressing subtype (HR=2.778, 95% CI: 1.585-4.867, P<0.001), TNBC (HR=3.832, 95% CI: 2.236-6.568, P<0.001), receipt of neoadjuvant chemotherapy (HR=1.992, 95% CI: 1.433-2.769, P<0.001), and visceral metastasis (HR=1.562, 95% CI: 1.122-2.173, P=0.008). No significant associations were found for histological grade II (P=0.351) or grade III (P=0.486), Luminal B subtype (P=0.403), surgical approach (P=0.425), adjuvant chemotherapy (P=0.331), or adjuvant radiotherapy (P=0.674).

Variables with P<0.05 in univariate analysis were entered into multivariate Cox regression. Results showed that independent risk factors for prOS included early recurrence (HR=2.578, 95% CI: 1.711-3.883, P<0.001), tumor size T3-T4 (HR=1.950, 95% CI: 1.321-2.878, P<0.001), lymph node positivity (HR=2.190, 95% CI: 1.512-3.172, P<0.001), HER2-overexpressing subtype (HR=2.347, 95% CI: 1.255-4.387, P=0.008), TNBC (HR=2.293, 95% CI: 1.250-4.205, P=0.007), receipt of neoadjuvant chemotherapy (HR=1.782, 95% CI: 1.277-2.488, P<0.001), and age younger than 35 years (HR=1.424, 95% CI: 1.012-2.003, P=0.043). Luminal B subtype (P=0.754) and visceral metastasis (P=0.956) were no longer statisti-

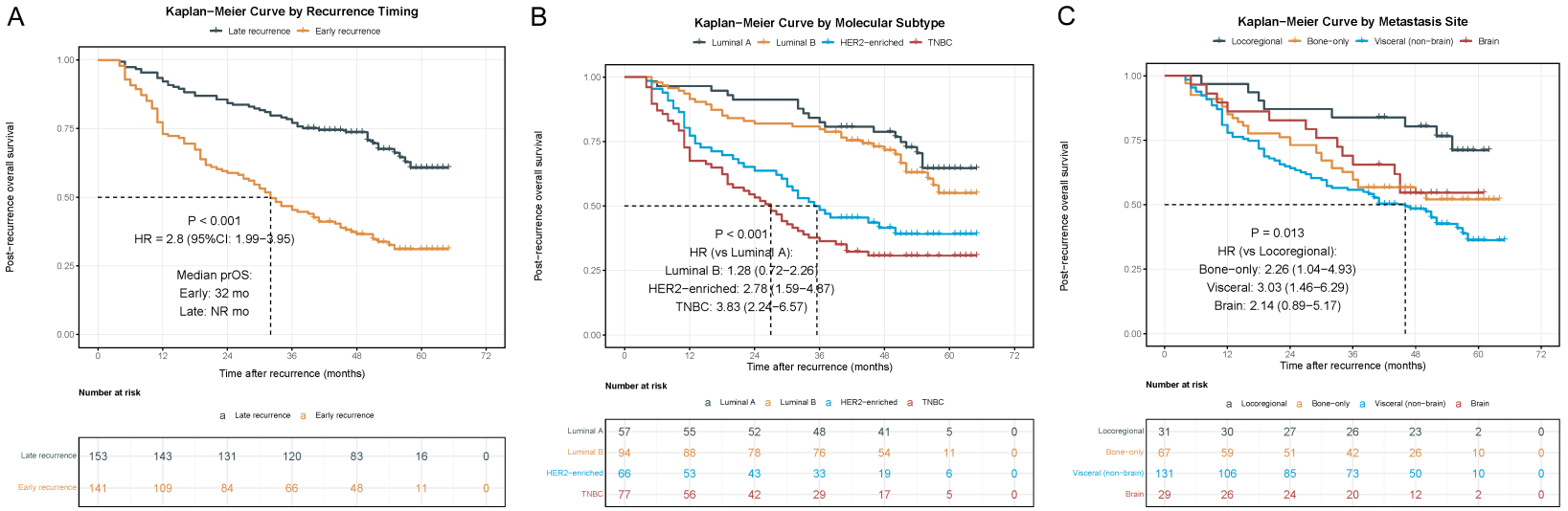

Figure 1. Kaplan-Meier survival curves for post-recurrence overall survival in the training cohort. A. Kaplan-Meier survival curves stratified by recurrence timing; B. Kaplan-Meier survival curves stratified by molecular subtype; C. Kaplan-Meier survival curves stratified by first recurrence site. Note: prOS, Post-recurrence overall survival; HR, Hazard ratio; CI, Confidence interval; TNBC, Triple-negative breast cancer; HER2, Human epidermal growth factor receptor 2; NR, Not reached.

Table 2. Variable coding and multicollinearity diagnosis for Cox regression model

Variable	Coding	VIF before adjustment	VIF after adjustment
Recurrence type	Yes =1, No =0	1.626	1.299
Age (years)	35-40=0, <35=1	1.088	1.057
Tumor size	T1-T2=0, T3-T4=1	1.156	1.085
Histological grade	Grade I =1, Grade II =2, Grade III =3	1.11	1.069
Lymph node status	N0 =0, N1-3=1	1.101	1.06
ER status	Positive =1, Negative =0	4.626	-
PR status	Positive =1, Negative =0	2.203	-
HER2 status	Positive =1, Negative =0	2.058	-
Ki-67 expression	Low (<30%)=0, High (≥30%)=1	1.513	-
Molecular subtype	Luminal A =1, Luminal B =2, HER2-overexpressing =3, TNBC=4	10.736	1.386
Surgical approach	Breast-conserving surgery =1, Mastectomy =0	1.157	1.068
Neoadjuvant chemotherapy	Yes =1, No =0	1.078	1.052
Adjuvant chemotherapy	Yes =1, No =0	1.07	1.049
Adjuvant radiotherapy	Yes =1, No =0	1.113	1.063
Adjuvant endocrine therapy	Yes =1, No =0	3.602	-
Distant metastasis	Yes =1, No =0	7.541	-
Locoregional recurrence	Yes =1, No =0	1.146	-
Bone metastasis	Yes =1, No =0	2.192	-
Liver metastasis	Yes =1, No =0	13.502	-
Lung metastasis	Yes =1, No =0	13.228	-
Brain metastasis	Yes =1, No =0	2.478	-
Visceral metastasis	Yes =1, No =0	14.856	1.25
Multi-organ metastasis	Yes =1, No =0	5.317	-

Note: VIF, Variance inflation factor; ER, Estrogen receptor; PR, Progesterone receptor; HER2, Human epidermal growth factor receptor 2; TNBC, Triple-negative breast cancer. "-" indicates variable was excluded from the final model due to multicollinearity.

cally significant after adjusting for other variables (see **Table 3**).

Comparison of prOS between early and late recurrence across molecular subtypes

To further explore how recurrence timing affects prognosis across molecular subtypes, we performed subgroup survival analyses.

In Luminal A patients, prOS differed significantly between early and late recurrence groups (P<0.001). Median prOS was 36 months in the early recurrence group versus not reached in the late recurrence group. Mortality risk was substantially higher in early recurrence patients (HR=8.67, 95% CI: 3.08-24.44).

In HER2-overexpressing patients, prOS also differed significantly between groups (P=0.011). Median prOS was 22 months for early recurrence and not reached for late recurrence. Early recurrence carried significantly elevated mortality risk (HR=2.23, 95% CI: 1.18-4.20).

In Luminal B patients, no significant prOS difference was observed between early and late recurrence groups (P=0.194). Neither group reached median prOS.

In TNBC patients, prOS did not differ significantly between groups (P=0.303). Median prOS was 26 months for early recurrence and 29 months for late recurrence (see **Figure 3**).

Comparison of prOS between early and late recurrence across first recurrence sites

To examine how recurrence timing influences prognosis according to first recurrence site, we conducted subgroup survival analyses. Recurrence sites were categorized by priority: brain metastasis, visceral metastasis (non-brain), bone-only metastasis, and locoregional recurrence.

In patients with locoregional recurrence, prOS differed significantly between early and late recurrence groups (P<0.001). Median prOS

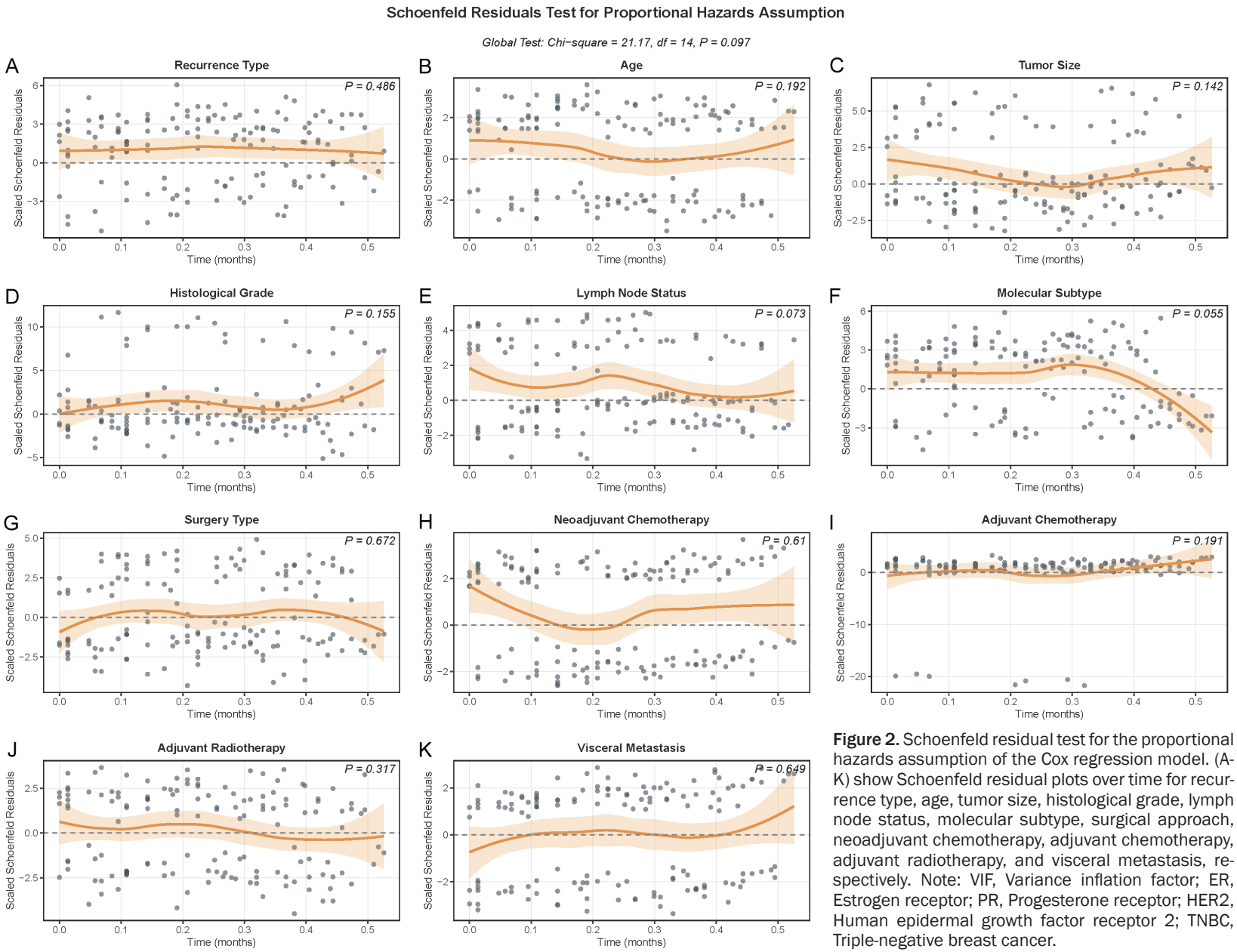

Figure 2. Schoenfeld residual test for the proportional hazards assumption of the Cox regression model. (A-K) show Schoenfeld residual plots over time for recurrence type, age, tumor size, histological grade, lymph node status, molecular subtype, surgical approach, neoadjuvant chemotherapy, adjuvant chemotherapy, adjuvant radiotherapy, and visceral metastasis, respectively. Note: VIF, Variance inflation factor; ER, Estrogen receptor; PR, Progesterone receptor; HER2, Human epidermal growth factor receptor 2; TNBC, Triple-negative breast cancer.

Table 3. Univariate and multivariate Cox regression analysis for post-recurrence overall survival in the training cohort

Variable	Univariate			Multivariate		
	β	*P* value	HR (95% CI)	β	*P* value	HR (95% CI)
Age (years)						
35-40 years			Reference			Reference
<35 years	0.453	0.008	1.574 (1.128-2.196)	0.353	0.043	1.424 (1.012-2.003)
Recurrence type						
Late recurrence			Reference			Reference
Early recurrence	1.03	<0.001	2.801 (1.985-3.953)	0.947	<0.001	2.578 (1.711-3.883)
Tumor size						
T1-T2			Reference			Reference
T3-T4	0.971	<0.001	2.640 (1.824-3.822)	0.668	<0.001	1.950 (1.321-2.878)
Histological grade						
Grade I			Reference			-
Grade II	-0.227	0.351	0.797 (0.494-1.285)			
Grade III	-0.172	0.486	0.842 (0.520-1.365)			
Lymph node status						
N0			Reference			Reference
N1-3	0.458	0.008	1.581 (1.128-2.216)	0.784	<0.001	2.190 (1.512-3.172)
Molecular subtype						
Luminal A			Reference			Reference
Luminal B	0.244	0.403	1.276 (0.720-2.261)	0.1	0.754	1.105 (0.591-2.066)
HER2-overexpressing	1.022	<0.001	2.778 (1.585-4.867)	0.853	0.008	2.347 (1.255-4.387)
TNBC	1.343	<0.001	3.832 (2.236-6.568)	0.83	0.007	2.293 (1.250-4.205)
Surgical approach						
Mastectomy			Reference			-
Breast-conserving surgery	0.135	0.425	1.145 (0.821-1.596)			
Neoadjuvant chemotherapy						
No			Reference			Reference
Yes	0.689	<0.001	1.992 (1.433-2.769)	0.578	<0.001	1.782 (1.277-2.488)
Adjuvant chemotherapy						
No			Reference			-
Yes	0.377	0.331	1.458 (0.682-3.115)			
Adjuvant radiotherapy						
No			Reference			-
Yes	0.071	0.674	1.073 (0.773-1.490)			
Visceral metastasis						
No			Reference			Reference
Yes	0.446	0.008	1.562 (1.122-2.173)	-0.010	0.956	0.990 (0.682-1.435)

Note: HR, Hazard ratio; CI, Confidence interval; HER2, Human epidermal growth factor receptor 2; TNBC, Triple-negative breast cancer.

was 26 months for early recurrence versus not reached for late recurrence. Early recurrence showed markedly elevated mortality risk (HR=7.94, 95% CI: 1.87-33.72).

In patients with visceral metastasis (non-brain), prOS also differed significantly (P<0.001). Median prOS was 28 months for early recurrence and 58 months for late recurrence. Early recurrence carried significantly higher mortality risk (HR=2.30, 95% CI: 1.41-3.74).

In brain metastasis patients, prOS showed a borderline difference between groups (P=

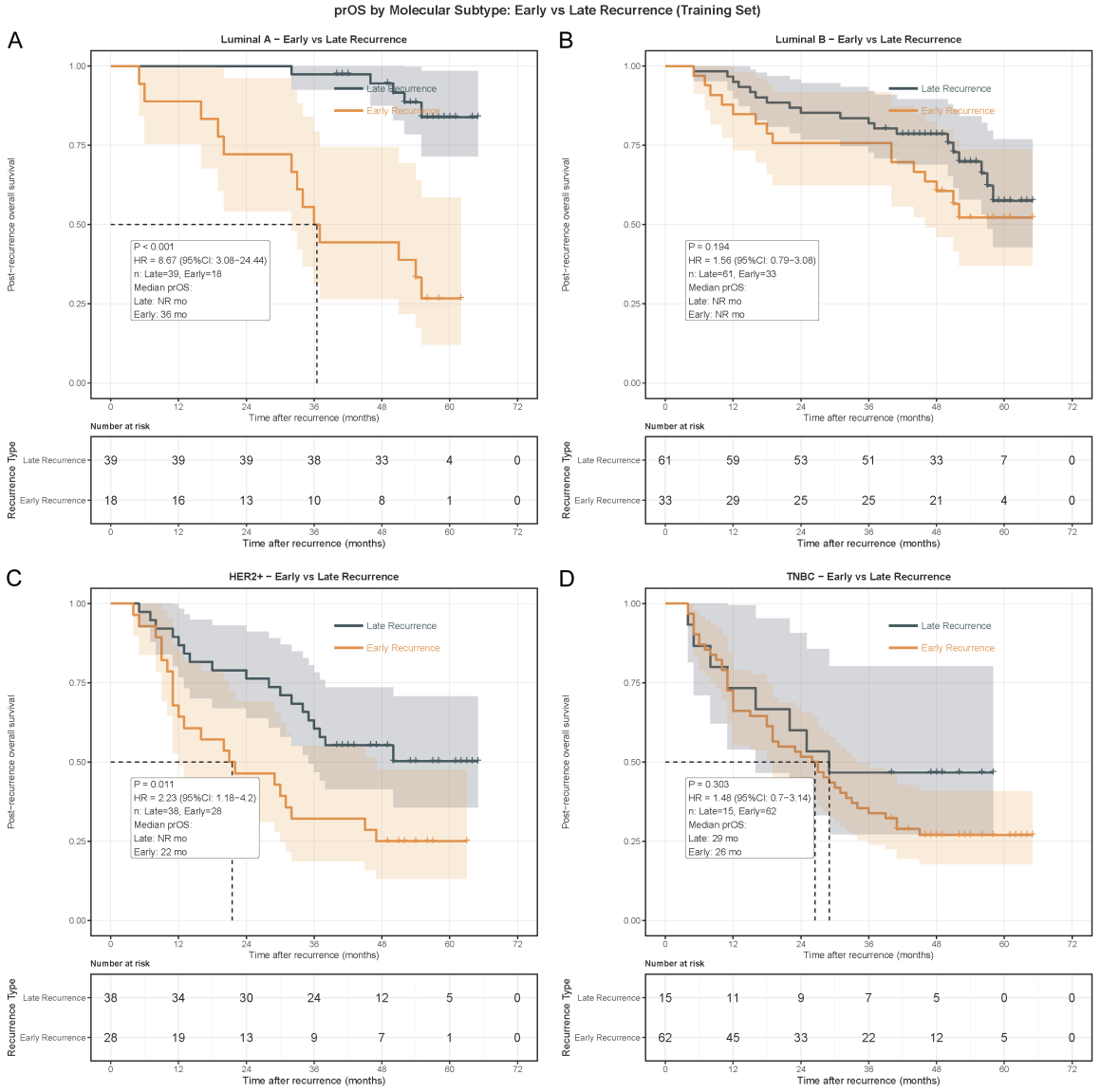

Figure 3. Kaplan-Meier survival curves comparing post-recurrence overall survival between early and late recurrence across molecular subtypes. A. Luminal A; B. Luminal B; C. HER2-overexpressing; D. Triple-negative breast cancer. Note: prOS, Post-recurrence overall survival; HR, Hazard ratio; CI, Confidence interval; HER2, Human epidermal growth factor receptor 2; TNBC, Triple-negative breast cancer; NR, Not reached.

0.049). Median prOS was 40 months for early recurrence versus not reached for late recurrence. However, the small number of late recurrence patients who developed brain metastasis (n=5) resulted in unstable HR estimates. These results should be interpreted with caution.

In bone-only metastasis patients, no significant prOS difference was observed between early and late recurrence groups (P=0.413). Median prOS was 40 months for early recurrence and not reached for late recurrence (see **Figure 4**).

Construction of the nomogram for predicting post-recurrence overall survival

Based on independent prognostic factors identified through multivariate Cox regression, we constructed a nomogram for predicting prOS in young breast cancer patients. Predictors included age, recurrence type, tumor size, lymph node status, molecular subtype, and neoadjuvant chemotherapy.

When employing this nomogram, first identify value for each predictor on its scale. Then proj-

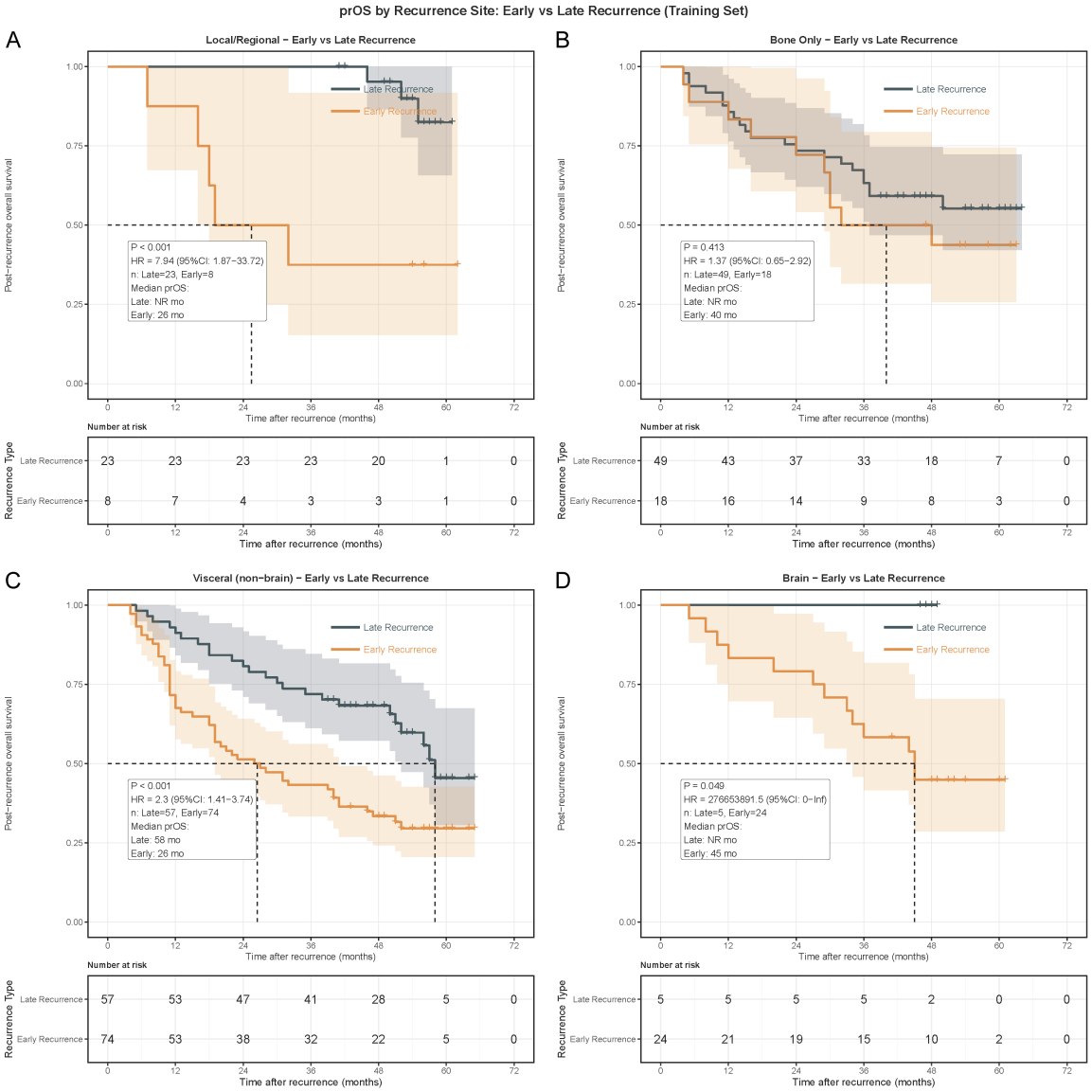

Figure 4. Kaplan-Meier survival curves comparing post-recurrence overall survival between early and late recurrence across first recurrence sites. A. Locoregional recurrence; B. Bone-only metastasis; C. Visceral metastasis (non-brain); D. Brain metastasis. Note: prOS, Post-recurrence overall survival; HR, Hazard ratio; CI, Confidence interval; NR, Not reached.

ect upwards to the 'Points' scale to derive that factor's score. Add all the factor scores to get aggregate points. Ultimately, extrapolate from the cumulative points to the survival probability scale to determine the forecasted probabilities of prOS at 1-year, 3-year, and 5-year intervals. The nomogram demonstrated that recurrence type and molecular subtype contributed the median number of total points and had the highest prognostic value for prOS. Patients with early recurrence, TNBC or HER2-overexpressing subtype, lymph node positivity (N1-3), large tumor size (T3-T4), receipt of neoadjuvant chemotherapy, and younger than 35 years accumulated higher total score, which corresponds to lower survival probability (**Figure 5**).

Internal validation of the prediction model in the training cohort

Time-dependent ROC curves assessed model discrimination. In the training cohort, AUC for predicting 1-year, 3-year, and 5-year prOS was 0.845 (95% CI: 0.783-0.906), 0.820 (95% CI: 0.773-0.868), and 0.649 (95% CI: 0.528-0.770), respectively. These values indicated

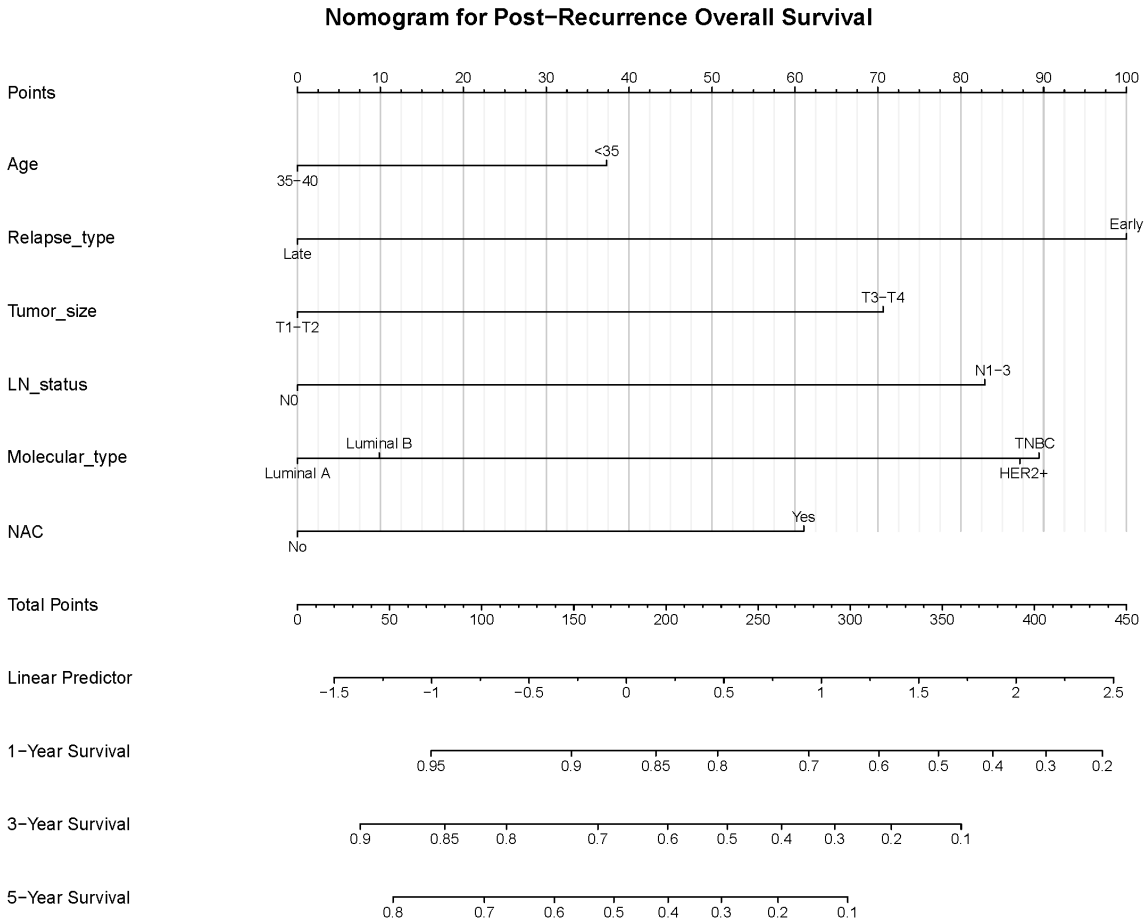

Figure 5. Nomogram for predicting post-recurrence overall survival in young breast cancer patients. The nomogram incorporates six predictors: age, recurrence type, tumor size, lymph node status, molecular subtype, and neoadjuvant chemotherapy. It is used to predict 1-year, 3-year, and 5-year post-recurrence overall survival probabilities. Note: prOS, Post-recurrence overall survival; TNBC, Triple-negative breast cancer; HER2, Human epidermal growth factor receptor 2.

good predictive ability for 1-year and 3-year prOS. The model C-index was 0.756 (95% CI: 0.719-0.794) (**Figure 6A**).

Bootstrap resampling (n=1000) for internal validation yielded a corrected C-index of 0.741 (95% CI: 0.734-0.757) with an optimism value of 0.015, suggesting minimal overfitting (**Figure 6B**).

Calibration curves assessed agreement between predicted and observed probabilities. The 3-year prOS calibration curve in the training cohort showed good concordance, with the calibration curve approaching the ideal diagonal line. Brier score was 0.171. The Hosmer-Lemeshow test was not statistically significant (P=0.121), indicating acceptable calibration.

Decision curve analysis showed that using the prediction model for clinical decision-making provided net benefit over “treat all” and “treat none” strategies within a certain threshold probability range (**Figure 6C**).

Using the training cohort median risk score (1.668) as cutoff, patients were stratified into high-risk and low-risk groups. Kaplan-Meier survival analysis revealed significant prOS differences between groups (P<0.001). Median prOS and HR for high-risk versus low-risk groups are shown in **Figure 6D** (see **Figure 6**).

External validation of the prediction model in the validation cohort

In the validation cohort, AUC for predicting 1-year, 3-year, and 5-year prOS was 0.720 (95%

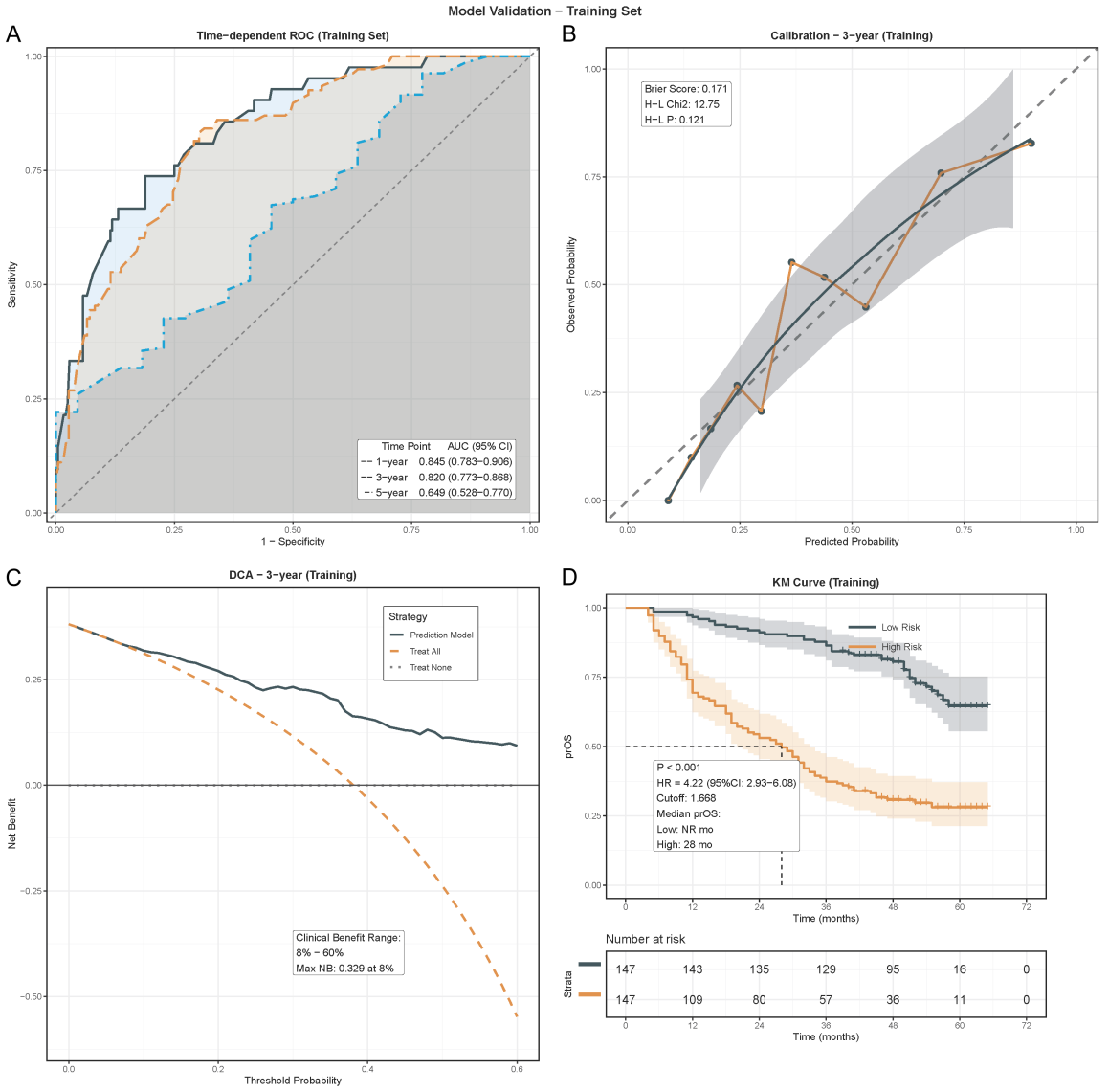

Figure 6. Internal validation of the prediction model in the training cohort. A. Time-dependent ROC curves; B. Calibration curve for 3-year post-recurrence overall survival; C. Decision curve analysis for 3-year post-recurrence overall survival; D. Kaplan-Meier survival curves based on risk score stratification. Note: ROC, Receiver operating characteristic; AUC, Area under the curve; CI, Confidence interval; C-index, Concordance index; DCA, Decision curve analysis; prOS, Post-recurrence overall survival; HR, Hazard ratio; NR, Not reached.

CI: 0.621-0.818), 0.725 (95% CI: 0.656-0.794), and 0.601 (95% CI: 0.463-0.740), respectively. These values indicated that the model retained acceptable predictive ability in an external population. The validation cohort C-index was 0.658 (95% CI: 0.604-0.711) (**Figure 7A**).

The 3-year prOS calibration curve showed a Brier score of 0.209. The Hosmer-Lemeshow test was statistically significant (P<0.001), suggesting that model calibration in the external population requires further optimization (**Figure 7B**).

Decision curve analysis demonstrated that the prediction model provided net benefit over "treat all" and "treat none" strategies within a certain threshold probability range (**Figure 7C**).

Using the same risk score cutoff as the training cohort (1.668), patients in the validation cohort were stratified by risk. Kaplan-Meier survival analysis showed significant prOS differences between high-risk and low-risk groups (P<0.001). Median prOS and HR for each risk group are shown in **Figure 7D** (see **Figure 7**).

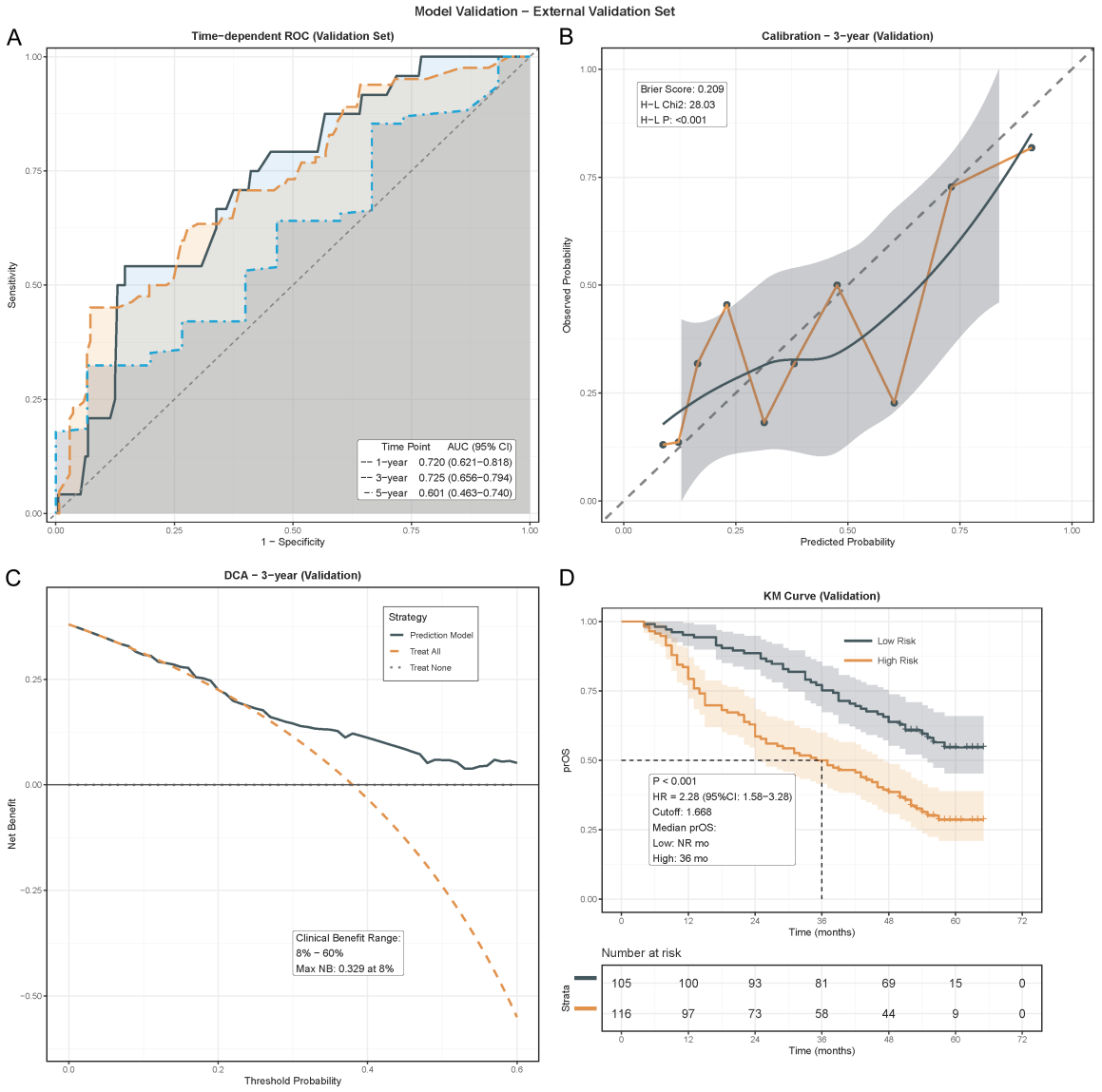

Figure 7. External validation of the prediction model in the validation cohort. A. Time-dependent ROC curves; B. Calibration curve for 3-year post-recurrence overall survival; C. Decision curve analysis for 3-year post-recurrence overall survival; D. Kaplan-Meier survival curves based on risk score stratification. Note: ROC, Receiver operating characteristic; AUC, Area under the curve; CI, Confidence interval; C-index, Concordance index; DCA, Decision curve analysis; prOS, Post-recurrence overall survival; HR, Hazard ratio; NR, Not reached.

Risk score distribution in training and validation cohorts

Individual risk scores were calculated from the Cox regression model. The training cohort median risk score (1.668) served as cutoff for stratifying patients into high-risk and low-risk groups.

Among 294 training cohort patients, risk scores ranged from 0 to 4.152, with mean 1.689 (standard deviation 0.943) and median 1.668 (interquartile range 1.023-2.312). The low-risk group contained 147 patients (50.0%), as did the high-risk group (147 patients, 50.0%). Density plots and histograms showed good separation between low-risk and high-risk group distributions.

Among 221 validation cohort patients, risk scores ranged from 0 to 3.821, with mean 1.710 (standard deviation 1.009) and median 1.770 (interquartile range 0.846-2.400). The low-risk group contained 105 patients (47.5%),

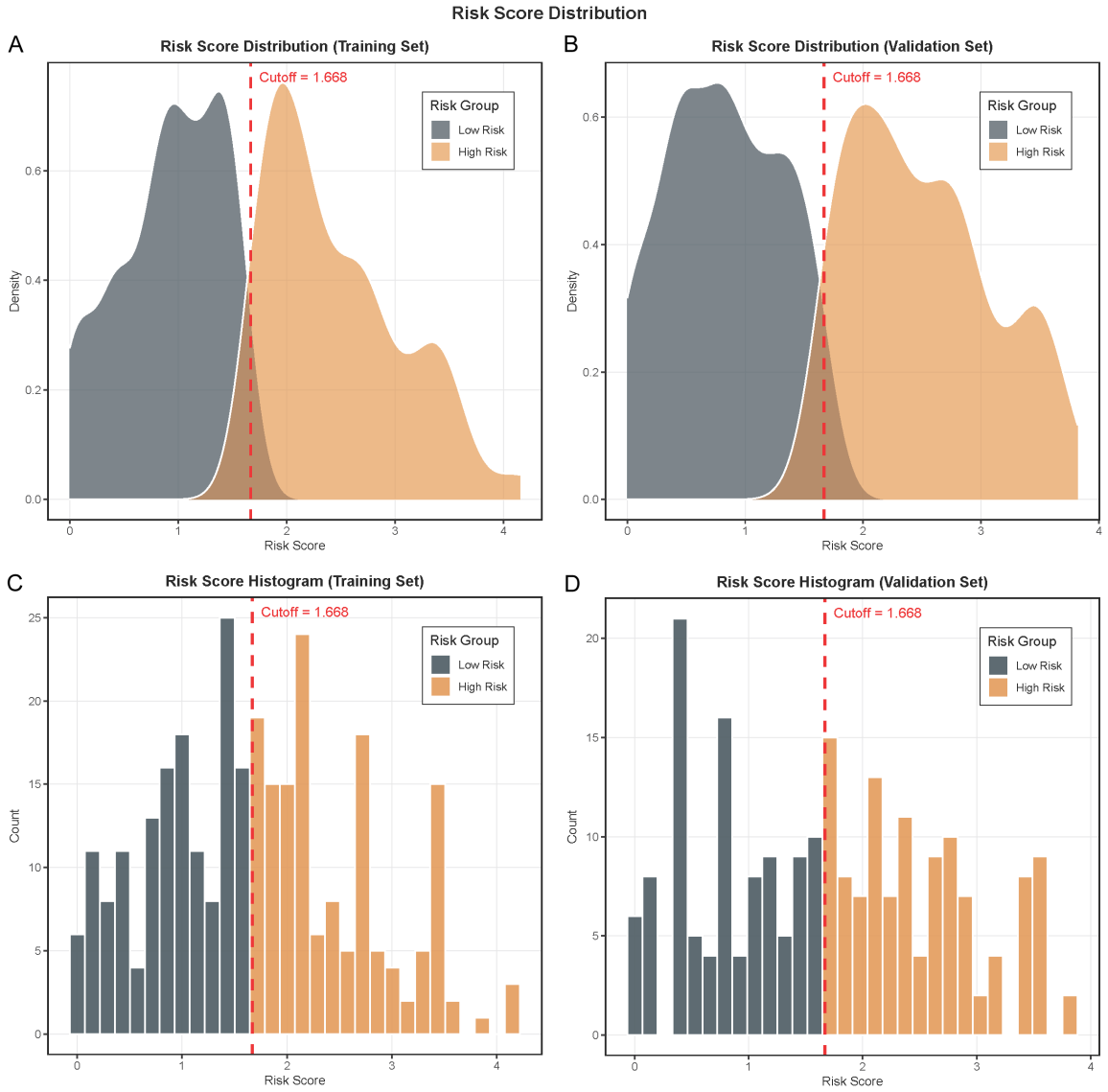

Figure 8. Risk score distribution in the training and validation cohorts. A. Density plot of risk scores in the training cohort; B. Density plot of risk scores in the validation cohort; C. Histogram of risk scores in the training cohort; D. Histogram of risk scores in the validation cohort. The red dashed line indicates the cutoff value for risk stratification (1.668). Note: IQR, Interquartile range; SD, Standard deviation.

while the high-risk group contained 116 patients (52.5%). Risk score distribution in the validation cohort was similar to that in the training cohort, demonstrating good consistency in distribution characteristics between cohorts (see **Figure 8**).

Discussion

Using data from 515 recurrent breast cancer patients aged 40 years or younger, this study systematically compared the clinicopathological characteristics of early and late recurrence and identified independent prognostic factors for post-recurrence overall survival (prOS) and built a nomogram. The findings have important clinical implications for personalized management and risk stratification of young breast cancer patients after recurrence.

Patients who experienced early recurrence had more aggressive tumor characteristics than patients with late recurrence, according to our study. Tumor size and grade was larger and more frequent lymph node positivity and ER/PR negativity. The observations are in agree-

ment with previous observations. As per the Carolina Breast Cancer Study, it was reported by Chollet-Hinton et al. [17] that breast cancers in young patients (less than 40 years) tended to be of a high-stage, hormone receptor-negative and basal-like subtype and had different biological characteristics than older patients. Kim et al. [7] similarly confirmed that younger patients were more likely to have HER2-positive or triple-negative subtypes at high stage, through a SEER database analysis.

As for the molecular subtype, we find that the majority of early recurrence cases had TNBC, at 43.97%, while the majority of late recurrence cases were Luminal. This pattern is seen at the characteristic recurrence time of different molecular subtypes. TNBC is a very aggressive disease and has a significantly higher early recurrence risk than other subtypes [18]. A research study which examines rapid recurrence in TNBC patients, those who have rapid recurrence (≤2 years) are more likely than those who do not have subsequent distant metastasis at first recurrence and greater visceral metastatic disease [19]. Conversely, hormone receptor-positive breast cancer has a late recurrence risk. According to Pan et al. [10], patients with ER-positive early breast cancer who completed 5 years of adjuvant endocrine therapy were still at risk of recurrence and death in the next 15 years.

Regarding recurrence and metastasis patterns, higher rates of distant metastasis, visceral metastasis, and multi-organ metastasis were found in the early recurrence group than in the late recurrence group. The more aggressive biological behavior of early recurring tumors likely explains this. Kennecke et al. [20] have demonstrated that distinct molecular subtypes have varying organ-specific metastatic potential. Breast cancers categorized as TNBC and HER2-positive metastasize more readily to visceral organs such as liver, lung, and brain while hormone receptor-positive breast cancers preference bone metastasis. In this study, the late recurrence group had higher locoregional recurrence, meanwhile, the rates of bone metastasis were not significantly different from the early recurrence group. This further suggests that patients with varying timing of recurrence have varying metastatic patterns.

According to our multivariate Cox regressions analysis, early recurrence was the best independent predictor of prOS (HR=2.578). The size of this effect is similar to those reported previously. Researchers Lee et al. [11] found that patients having early recurrence showed significantly higher mortality risk than those with late occurrence. According to Narod et al. [21], patients who have local recurrence occurring within 2 years have worse outcomes compared to those with recurrence occurring after 2 years (adjusted HR=0.38). When tumors were found to recur shortly after treatment, it suggests that they have a greater capacity for proliferation, higher heterogeneity, and more resistant forms toward multiple anti-cancer drugs. All these biological characteristics lead to a poorer prognosis [22].

Molecular subtype had significant impact on pre-OS. The findings our study indicated that both HER2-overexpressing subtype (HR=2.347) and TNBC (HR=2.293) independently predicted prOS. Triple-negative breast cancer (TNBC) does not have effective targeted therapy options. Despite the advent of neoadjuvant therapy, TNBC remains prognostically worse than other subtypes, by a statistically significant margin [23]. The outcomes of HER2-positive breast cancer have improved considerably with anti-HER2 targeted therapy. However, patient-reported overall survival or prOS is still suboptimal after recurrence. This indicates a need for improved later-line treatment strategies [24].

In our study, independent risk factors of prOS included tumor burden information of the T3-T4 stage and lymph node positivity. Prior research established significant correlations between breast cancer recurrence and death, and initial diagnosis of tumor size and lymph node involvement [25]. Even with defined pathological complete response (pCR), initial clinical lymph nodes positivity (cN+) is associated with higher recurrence risk [26, 27]. The biological characteristics of the primary tumor appear to impact post-recurrence survival.

In our study, receiving neoadjuvant chemotherapy was an independent risk factor of prOS (HR=1.782). Despite being paradoxical at face value, meticulous interpretation is required. Those diagnosed with more advanced cancer of a more aggressive type are usually given

neoadjuvant chemotherapy. We could not include pCR status in our study. In a meta-analysis carried out by Spring et al., there was an observed improvement in both event-free survival (EFS) and overall survival (OS) among patients who achieved pCR after neoadjuvant chemotherapy. The correlation between pCR and improved survival outcomes was particularly strong in patients with TNBC and HER2-positive disease. Furthermore, we could not characterise the true expression of neoadjuvant treatment response on prOS due to the lack of pCR status and post-recurrence treatment data in our study. Consequently, the interpretation of this result must be cautious. It indicates an effect and not an actual cause.

Our study found being less than 35 years old to be an independent risk of prOS (HR=1.424). It has to do with the unique tumor biology of young breast cancer. Research indicates that younger patients' tumors show higher proliferative activity, greater stem cell features, and enriched immunosuppressive signaling [28]. In a broad cohort study from Denmark, Pedersen et al. [12] also demonstrated that diagnosis under age 40 was an independent risk factor for late recurrence. This implies that the likelihood of recurrence among young patients is increased even 10 years after diagnosis. This outcome supports the view that young breast cancer is a biological entity.

According to molecular subtype specific offspring that are early and late recurrence subgroup analysis statistics differences were observed in prOS. In patients with Luminal A and HER2-overexpressing cancer, the prOS was significantly worse for the early recurrence group compared with the late recurrence group. Nonetheless, no significant differences were found in Luminal B and TNBC patients. The clinical implications of this finding expandable.

For Luminal A patients, who are generally considered to have favorable prognosis, our study showed that once early recurrence occurs, prOS drops dramatically. Median prOS was only 36 months, with mortality risk nearly 9-fold higher than the late recurrence group. This indicates that even for low-risk subtypes, early recurrence signals poor prognosis. Lee et al. [11] similarly found that in the Luminal B subtype, late recurrence was an independent prognostic factor for overall survival, whereas no significant difference was observed in TNBC and HER2 subtypes.

The lack of significant prOS difference between early and late recurrence in TNBC patients in our study may relate to the inherently poor baseline prognosis of TNBC. Regardless of whether recurrence is early or late, TNBC patients have relatively short prOS, which may obscure differences between groups. This negative result may also be limited by insufficient subgroup sample size, reducing statistical power. Validation in larger samples is needed.

Subgroup analysis by first recurrence site showed that prOS differences between early and late recurrence were significant in patients with locoregional recurrence and visceral metastasis (non-brain), but not significant in bone-only metastasis patients. This may relate to the relatively indolent biological behavior of bone metastasis. Previous studies suggest that breast cancer patients with bone-only metastasis have substantially longer median survival than those with visceral metastasis [29]. The bone microenvironment may provide conditions favorable for maintaining tumor cell "dormancy", thereby delaying disease progression [30].

Based on independent prognostic factors from multivariate Cox regression, we constructed a nomogram for predicting prOS. In the training cohort, the model achieved a C-index of 0.756, with 1-year and 3-year AUC of 0.845 and 0.820, respectively. These values indicate good discrimination. After Bootstrap internal validation, the C-index was 0.741, suggesting minimal overfitting. In the validation cohort, the C-index was 0.658. Although somewhat lower than the training cohort, the model demonstrated acceptable generalizability.

Our nomogram possesses certain benefits over previously developed prediction models. CTS5 or Clinical Treatment Score post-5 years is a commonly used late recurrence risk prediction tool, which only applies to ER-positive patients who remain recurrence free after 5 years and predicts subsequent recurrence risk [31]. Genomic tests e.g. Oncotype DX can provide finer molecular information. However, they are costly and not implementable in the resource-poor settings [32]. Our nomogram incorporates easily accessible clinical characteristics for easy implementation in clinical practice.

It can be a useful tool to predict prOS in young patients with recurrent breast cancer.

Nonetheless, our model exhibited suboptimal calibration within the validation cohort (H-L test P<0.001). This means that the predicted and observed probabilities are not aligned. The H-L test should, however, be noted as being sensitive to sample size, showing important biases in medium-sized validation cohorts [33]. Different characteristics of the patients treated and differences in treatment between centres could impact external applicability. Future confirmation and optimization in broader multicenter prospective cohorts are essential.

Based on risk stratification analysis using risk scores, high-risk and low-risk prOS differed significantly in both training and validation cohorts. The information from stratifying risk can guide clinical decision making. High-risk patients may need more intensive treatment and closer follow-up monitoring, whereas low-risk patients may consider de-escalation of appropriate treatment to reduce toxicity.

This study is not without limitations. Firstly, there may be selection bias and information bias due to retrospective study. Secondly, we were unable to incorporate post-recurrence treatment information, as treatment response following recurrence significantly impacts prOS. The fact that pCR status was not included may have affected the finding that neo-adjuvant chemotherapy is a risk factor. The cutoff of Ki-67 remains controversial. We utilized a cutoff of 30%, but varying studies and guidelines recommend different cutoffs [34]. Fifth, genomic test results like Oncotype DX scores were not included. These molecular markers could provide additional prognostic information. Sixth, because the validation cohort had a relatively small sample size and was from one center, geographic representativeness was limited. The follow up duration was quite short; some patients' long-term outcomes were not observed. Future research could consider several potential directions. First, further inclusion of post-recurrence treatment response, circulating tumor DNA (ctDNA), and other dynamic indicators could lead to better prognostic prediction models [35]. Second, we can conduct prospective multicenter validation studies to further assess the external applicability. Next, genomic features like Oncotype DX scores and MammaPrint could also be added to assess clinicopathological characteristics and molecular markers. Finally, one could consider developing online calculators or mobile apps for rapid clinical risk assessment or clinical application.

Conclusions

In conclusion, this study revealed that early recurrence is a significant predictor of poor post-recurrence survival in young breast cancer patients 40 years of age and younger. Patients with early recurrence have more aggressive tumor characteristics and a greater chance of visceral metastasis. A nomogram that includes recurrence time points, tumor burden, and molecular type is a reasonable predictor of prOS in young recurrent breast cancer patients and has acceptable generalizability to an external population. This model may be useful for clinical risk stratification and individualized management to define high-risk patients and establish treatment and follow-up strategies.

Disclosure of conflict of interest

None.

Address correspondence to: Yun Cai, Department of Traditional Chinese Medicine (TCM), The First Affiliated Hospital of Xi'an Jiaotong University, No. 277, Yanta West Road, Xi'an 710061, Shaanxi, China. E-mail: yhyqiu01@163.com

References

[1] Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A and Bray F. 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.

[2] Sun K, Zhang B, Lei S, Zheng R, Liang X, Li L, Feng X, Zhang S, Zeng H, Yao Y, Ma P, Wang S, Chen R, Han B, Wei W and He J. Incidence, mortality, and disability-adjusted life years of female breast cancer in China, 2022. Chin Med J (Engl) 2024; 137: 2429-2436.

[3] Xu H and Xu B. Epidemiology, early detection, and management of breast cancer in China: a comprehensive review. Chin J Cancer Res 2025; 37: 882-899.

[4] Sun K, Lei L, Zheng R, Zhang S, Zeng H, Wang S, Li L, Chen R, Han B, Peng J, Wei W and He J. Trends in incidence rates, mortality rates, and age-period-cohort effects of female breast

cancer - China, 2003-2017. China CDC Wkly 2023; 5: 340-346.
[5] Assi HA, Khoury KE, Dbouk H, Khalil LE, Mouhieddine TH and El Saghir NS. Epidemiology and prognosis of breast cancer in young women. J Thorac Dis 2013; 5 Suppl 1: S2-8.
[6] Duffy SW, Vulkan D, Cuckle H, Parmar D, Sheikh S, Smith RA, Evans A, Blyuss O, Johns L, Ellis IO, Myles J, Sasieni PD and Moss SM. Effect of mammographic screening from age 40 years on breast cancer mortality (UK Age trial): final results of a randomised, controlled trial. Lancet Oncol 2020; 21: 1165-1172.
[7] Kim HJ, Kim S, Freedman RA and Partridge AH. The impact of young age at diagnosis (age <40 years) on prognosis varies by breast cancer subtype: A U.S. SEER database analysis. Breast 2022; 61: 77-83.
[8] Zhu JW, Charkhchi P, Adekunte S and Akbari MR. What is known about breast cancer in young women? Cancers (Basel) 2023; 15: 1917.
[9] Courtney D, Davey MG, Moloney BM, Barry MK, Sweeney K, McLaughlin RP, Malone CM, Lowery AJ and Kerin MJ. Breast cancer recurrence: factors impacting occurrence and survival. Ir J Med Sci 2022; 191: 2501-2510.
[10] Pan H, Gray R, Braybrooke J, Davies C, Taylor C, McGale P, Peto R, Pritchard KI, Bergh J, Dowsett M and Hayes DF. 20-year risks of breast-cancer recurrence after stopping endocrine therapy at 5 years. N Engl J Med 2017; 377: 1836-1846.
[11] Lee YJ, Jung SP, Bae JW, Yang SM, You JY and Bae SY. Prognosis according to the timing of recurrence in breast cancer. Ann Surg Treat Res 2023; 104: 1-9.
[12] Pedersen RN, Esen B, Mellemkjær L, Christiansen P, Ejlertsen B, Lash TL, Nørgaard M and Cronin-Fenton D. The incidence of breast cancer recurrence 10-32 years after primary diagnosis. J Natl Cancer Inst 2022; 114: 391-399.
[13] Lin S, Mo H, Li Y, Guan X, Chen Y, Wang Z, Yuan P, Wang J, Luo Y, Fan Y, Cai R, Li Q, Chen S, Zhang P, Li Q, Ma F and Xu B. Development and validation of a nomogram for predicting survival of advanced breast cancer patients in China. Breast 2020; 53: 172-180.
[14] Balachandran VP, Gonen M, Smith JJ and DeMatteo RP. Nomograms in oncology: more than meets the eye. Lancet Oncol 2015; 16: e173-180.
[15] Chen X, Fan Y and Xu B. Distinct characteristics and metastatic behaviors of late recurrence in patients with hormone receptor-positive/human epidermal growth factor receptor 2-negative breast cancer: a single institute experience of more than 10 years. Clin Breast Cancer 2018; 18: e1353-e1360.
[16] Lu WL, Jansen L, Post WJ, Bonnema J, Van de Velde JC and De Bock GH. Impact on survival of early detection of isolated breast recurrences after the primary treatment for breast cancer: a meta-analysis. Breast Cancer Res Treat 2009; 114: 403-412.
[17] Chollet-Hinton L, Anders CK, Tse CK, Bell MB, Yang YC, Carey LA, Olshan AF and Troester MA. Breast cancer biologic and etiologic heterogeneity by young age and menopausal status in the carolina breast cancer study: a case-control study. Breast Cancer Res 2016; 18: 79.
[18] Dent R, Trudeau M, Pritchard KI, Hanna WM, Kahn HK, Sawka CA, Lickley LA, Rawlinson E, Sun P and Narod SA. Triple-negative breast cancer: clinical features and patterns of recurrence. Clin Cancer Res 2007; 13: 4429-4434.
[19] Cai SL, Liu JJ, Liu YX, Yu SH, Liu X, Lin XQ, Chen HD, Fang X, Ma T, Li YQ, Li Y, Li CY, Zhang S, Chen XG, Guo XJ and Zhang J. Characteristics of recurrence, predictors for relapse and prognosis of rapid relapse triple-negative breast cancer. Front Oncol 2023; 13: 1119611.
[20] Kennecke H, Yerushalmi R, Woods R, Cheang MC, Voduc D, Speers CH, Nielsen TO and Gelmon K. Metastatic behavior of breast cancer subtypes. J Clin Oncol 2010; 28: 3271-3277.
[21] Dent R, Valentini A, Hanna W, Rawlinson E, Rakovitch E, Sun P and Narod SA. Factors associated with breast cancer mortality after local recurrence. Curr Oncol 2014; 21: e418-425.
[22] Kvokačková B, Remšík J, Jolly MK and Souček K. Phenotypic heterogeneity of triple-negative breast cancer mediated by epithelial-mesenchymal plasticity. Cancers (Basel) 2021; 13: 2188.
[23] Bianchini G, De Angelis C, Licata L and Gianni L. Treatment landscape of triple-negative breast cancer - expanded options, evolving needs. Nat Rev Clin Oncol 2022; 19: 91-113.
[24] Huober J, van Mackelenbergh M, Schneeweiss A, Seither F, Blohmer JU, Denkert C, Tesch H, Hanusch C, Salat C, Rhiem K, Solbach C, Fasching PA, Jackisch C, Reinisch M, Lederer B, Mehta K, Link T, Nekljudova V, Loibl S and Untch M. Identifying breast cancer patients at risk of relapse despite pathological complete response after neoadjuvant therapy. NPJ Breast Cancer 2023; 9: 23.
[25] Colleoni M, Sun Z, Price KN, Karlsson P, Forbes JF, Thürlimann B, Gianni L, Castiglione M, Gelber RD, Coates AS and Goldhirsch A. Annual hazard rates of recurrence for breast cancer during 24 years of follow-up: results from the international breast cancer study group trials I to V. J Clin Oncol 2016; 34: 927-935.
[26] von Minckwitz G, Untch M, Blohmer JU, Costa SD, Eidtmann H, Fasching PA, Gerber B, Eiermann W, Hilfrich J, Huober J, Jackisch C,

Kaufmann M, Konecny GE, Denkert C, Nekljudova V, Mehta K and Loibl S. Definition and impact of pathologic complete response on prognosis after neoadjuvant chemotherapy in various intrinsic breast cancer subtypes. J Clin Oncol 2012; 30: 1796-1804.

[27] Spring LM, Fell G, Arfe A, Sharma C, Greenup R, Reynolds KL, Smith BL, Alexander B, Moy B, Isakoff SJ, Parmigiani G, Trippa L and Bardia A. 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.

[28] Azim HA Jr and Partridge AH. Biology of breast cancer in young women. Breast Cancer Res 2014; 16: 427.

[29] Van Mechelen M, Van Herck A, Punie K, Nevelsteen I, Smeets A, Neven P, Weltens C, Han S, Vanderstichele A, Floris G, Lobelle JP and Wildiers H. Behavior of metastatic breast cancer according to subtype. Breast Cancer Res Treat 2020; 181: 115-125.

[30] Coleman RE, Croucher PI, Padhani AR, Clézardin P, Chow E, Fallon M, Guise T, Colangeli S, Capanna R and Costa L. Bone metastases. Nat Rev Dis Primers 2020; 6: 83.

[31] Dowsett M, Sestak I, Regan MM, Dodson A, Viale G, Thürlimann B, Colleoni M and Cuzick J. Integration of clinical variables for the prediction of late distant recurrence in patients with estrogen receptor-positive breast cancer treated with 5 years of endocrine therapy: CTS5. J Clin Oncol 2018; 36: 1941-1948.

[32] Wang C, Xu Y, Lin Y, Zhou Y, Mao F, Zhang X, Shen S, Zhang Y and Sun Q. Comparison of CTS5 risk model and 21-gene recurrence score assay in large-scale breast cancer population and combination of CTS5 and recurrence score to develop a novel nomogram for prognosis prediction. Breast 2022; 63: 61-70.

[33] Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, Pencina MJ and Kattan MW. Assessing the performance of prediction models: a framework for traditional and novel measures. Epidemiology 2010; 21: 128-138.

[34] Nielsen TO, Leung SCY, Rimm DL, Dodson A, Acs B, Badve S, Denkert C, Ellis MJ, Fineberg S, Flowers M, Kreipe HH, Laenkholm AV, Pan H, Penault-Llorca FM, Polley MY, Salgado R, Smith IE, Sugie T, Bartlett JMS, McShane LM, Dowsett M and Hayes DF. Assessment of Ki67 in breast cancer: updated recommendations from the international Ki67 in breast cancer working group. J Natl Cancer Inst 2021; 113: 808-819.

[35] Magbanua MJM, Swigart LB, Wu HT, Hirst GL, Yau C, Wolf DM, Tin A, Salari R, Shchegrova S, Pawar H, Delson AL, DeMichele A, Liu MC, Chien AJ, Tripathy D, Asare S, Lin CJ, Billings P, Aleshin A, Sethi H, Louie M, Zimmermann B, Esserman LJ and van't Veer LJ. Circulating tumor DNA in neoadjuvant-treated breast cancer reflects response and survival. Ann Oncol 2021; 32: 229-239.

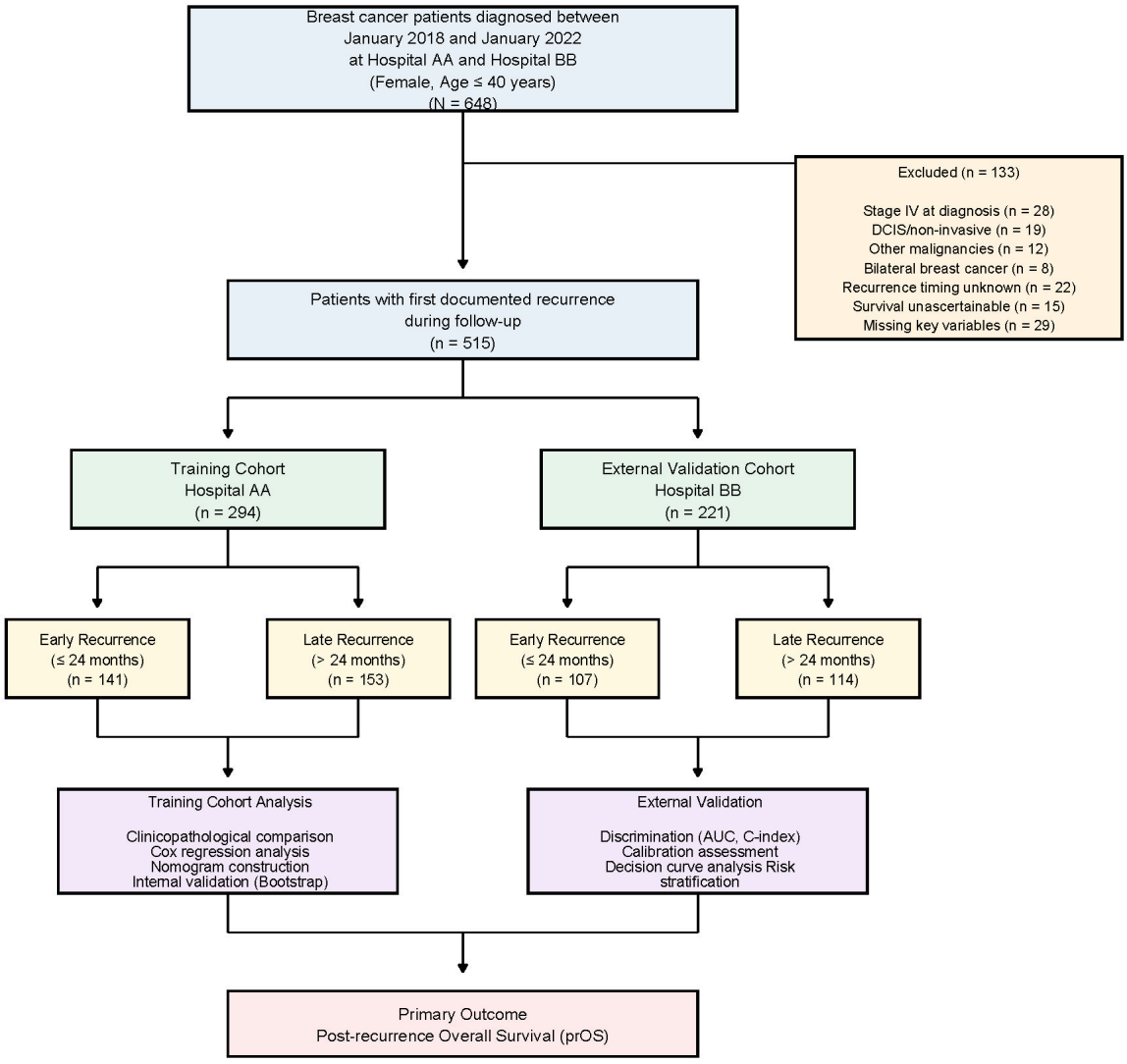

Figure S1. Flowchart of the study.

Table S1. Comparison of baseline characteristics between training and validation cohorts

Characteristic	Total	Training Cohort (n=294)	validation cohort (n=221)	Statistic	*P* value
Recurrence type				0.011	0.918
Late	267 (51.84%)	153 (52.04%)	114 (51.58%)		
Early	248 (48.16%)	141 (47.96%)	107 (48.42%)		
Age (years)				0.793	0.373
36-40 years	240 (46.60%)	142 (48.30%)	98 (44.34%)		
≤35 years	275 (53.40%)	152 (51.70%)	123 (55.66%)		
Tumor size				0.174	0.677
T1-T2	427 (82.91%)	242 (82.31%)	185 (83.71%)		
T3-T4	88 (17.09%)	52 (17.69%)	36 (16.29%)		
Histological grade				0.200	0.905
Grade I	73 (14.17%)	40 (13.61%)	33 (14.93%)		
Grade II	232 (45.05%)	134 (45.58%)	98 (44.34%)		
Grade III	210 (40.78%)	120 (40.82%)	90 (40.72%)		
Lymph node status				3.319	0.068
N1-3	335 (65.05%)	201 (68.37%)	134 (60.63%)		
N0	180 (34.95%)	93 (31.63%)	87 (39.37%)		
ER status				1.150	0.283
Negative	275 (53.40%)	163 (55.44%)	112 (50.68%)		
Positive	240 (46.60%)	131 (44.56%)	109 (49.32%)		
PR status				0.134	0.714
Negative	338 (65.63%)	191 (64.97%)	147 (66.52%)		
Positive	177 (34.37%)	103 (35.03%)	74 (33.48%)		
HER2 status				0.002	0.964
Negative	328 (63.69%)	187 (63.61%)	141 (63.80%)		
Positive	187 (36.31%)	107 (36.39%)	80 (36.20%)		
Ki-67 expression				0.165	0.684
Low (<30%)	286 (55.53%)	161 (54.76%)	125 (56.56%)		
High (≥30%)	229 (44.47%)	133 (45.24%)	96 (43.44%)		
Molecular subtype				1.295	0.730
Luminal A	102 (19.81%)	57 (19.39%)	45 (20.36%)		
Luminal B	171 (33.20%)	94 (31.97%)	77 (34.84%)		
HER2-overexpressing	107 (20.78%)	66 (22.45%)	41 (18.55%)		
TNBC	135 (26.21%)	77 (26.19%)	58 (26.24%)		
Surgical approach				0.171	0.679
Mastectomy	321 (62.33%)	181 (61.56%)	140 (63.35%)		
Breast-conserving surgery	194 (37.67%)	113 (38.44%)	81 (36.65%)		
Neoadjuvant chemotherapy				0.552	0.458
No	298 (57.86%)	166 (56.46%)	132 (59.73%)		
Yes	217 (42.14%)	128 (43.54%)	89 (40.27%)		
Adjuvant chemotherapy				0.037	0.847
No	36 (6.99%)	20 (6.80%)	16 (7.24%)		
Yes	479 (93.01%)	274 (93.20%)	205 (92.76%)		
Adjuvant radiotherapy				0.023	0.880
No	235 (45.63%)	135 (45.92%)	100 (45.25%)		
Yes	280 (54.37%)	159 (54.08%)	121 (54.75%)		

Adjuvant endocrine therapy				0.214	0.644
No	281 (54.56%)	163 (55.44%)	118 (53.39%)		
Yes	234 (45.44%)	131 (44.56%)	103 (46.61%)		
Distant metastasis				0.451	0.502
No	123 (23.88%)	67 (22.79%)	56 (25.34%)		
Yes	392 (76.12%)	227 (77.21%)	165 (74.66%)		
Locoregional recurrence				2.481	0.115
No	306 (59.42%)	166 (56.46%)	140 (63.35%)		
Yes	209 (40.58%)	128 (43.54%)	81 (36.65%)		
Bone metastasis				0.292	0.589
No	310 (60.19%)	174 (59.18%)	136 (61.54%)		
Yes	205 (39.81%)	120 (40.82%)	85 (38.46%)		
Liver metastasis				0.889	0.346
No	354 (68.74%)	207 (70.41%)	147 (66.52%)		
Yes	161 (31.26%)	87 (29.59%)	74 (33.48%)		
Lung metastasis				0.002	0.966
No	378 (73.40%)	216 (73.47%)	162 (73.30%)		
Yes	137 (26.60%)	78 (26.53%)	59 (26.70%)		
Brain metastasis				0.282	0.595
No	461 (89.51%)	265 (90.14%)	196 (88.69%)		
Yes	54 (10.49%)	29 (9.86%)	25 (11.31%)		
Visceral metastasis				0.000	1.000
No	261 (50.68%)	149 (50.68%)	112 (50.68%)		
Yes	254 (49.32%)	145 (49.32%)	109 (49.32%)		
Multi-organ metastasis				0.911	0.340
No	451 (87.57%)	261 (88.78%)	190 (85.97%)		
Yes	64 (12.43%)	33 (11.22%)	31 (14.03%)		

Note: ER, Estrogen receptor; PR, Progesterone receptor; HER2, Human epidermal growth factor receptor 2; TNBC, Triple-negative breast cancer.