The Breast 86 (2026) 104738

Contents lists available at ScienceDirect

The Breast

journal homepage: www.journals.elsevier.com/the-breast

Late distant recurrence prediction model in premenopausal women with ER-positive/HER 2-negative breast cancer: A multicenter retrospective study

Dong Seung Shin [a,1,2], Janghee Lee [b,1], Eunhye Kang [c,1], Dasom Noh [d,1], Jong-Ho Cheun [e], Jun-Hee Lee [f], Yeongyeong Son [d], Soong June Bae [g], Sunyoung Kwon [d,h,i], Han-Byoel Lee [c,j,k,*], Jai Min Ryu [a,**], Sung Gwe Ahn [g,l,***]

[a] Division of Breast Surgery, Department of Surgery, Samsung Medical Center, Sungkyunkwan University of Medicine, Seoul, Republic of Korea
[b] Department of Surgery, Ewha Womans University Mokdong Hospital, Ewha Womans University College of Medicine, Seoul, Republic of Korea
[c] Department of Surgery, Seoul National University Hospital, Seoul National University College of Medicine, Seoul, Republic of Korea
[d] Department of Information Convergence Engineering, College of Information and Biomedical Engineering, Pusan National University, Busan, Republic of Korea
[e] Department of Surgery, Seoul Metropolitan Government–Seoul National University Boramae Medical Center, Seoul, Republic of Korea
[f] Department of Surgery, Soonchunhyang University College of Medicine, Soonchunhyang University Hospital, Seoul, Republic of Korea
[g] Department of Surgery, Gangnam Severance Hospital, Yonsei University College of Medicine, Seoul, Republic of Korea
[h] School of Biomedical Convergence Engineering, College of Information and Biomedical Engineering, Pusan National University, Yangsan, Republic of Korea
[i] Center for Artificial Intelligence Research, Pusan National University, Busan, Republic of Korea
[j] Cancer Research Institute, Seoul National University, Seoul, Republic of Korea
[k] Biomedical Research Institute, Seoul National University Hospital, Seoul, Republic of Korea
[l] Institute for Breast Cancer Precision Medicine, Gangnam Severance Hospital, Yonsei University College of Medicine, Seoul, Republic of Korea

ARTICLE INFO

Keywords:
Breast cancer
Late distant recurrence
Machine-learning prediction model
Premenopausal
Extended endocrine therapy

ABSTRACT

Background: Late distant recurrence (DR) remains a significant challenge in estrogen receptor (ER)–positive/Human Epidermal Growth Factor Receptor 2 (HER2)–negative breast cancer, especially in premenopausal patients. This study aimed to develop a machine-learning model predicting late DR risk in premenopausal patients and to assess the clinical benefit of extended endocrine therapy (ET) according to risk stratification.

Methods: This retrospective multicenter study included patients aged ≤45 years with ER–positive/HER2–-negative breast cancer who underwent surgery between January 2000 and December 2011. This study was designed as a landmark analysis, with the effective baseline set at 5 years after surgery. Eligible patients had five to 10 years of follow-up and received adjuvant ET for at least two years. The primary outcome was late DR, defined as distant metastasis occurring between five and 10 years after surgery.

Results: Among 1701 included patients (median age, 41 years), late DR occurred in 108 patients (6.3%). A machine-learning model using eight clinicopathologic variables demonstrated strong predictive performance (AUC = 0.78). Patients classified as high-risk by the model exhibited significantly worse late DMFS compared to low-risk patients (HR,7.36; 95% CI,4.43–12.20; $P < 0.001$). Among high-risk patients, those who received extended ET had significantly improved late DMFS compared to those who did not (HR,0.32; 95% CI,0.18–0.55; $P < 0.001$). In low-risk patients, extended ET was not associated with a statistically significant benefit (HR,0.45; 95% CI,0.16–1.22; $P = 0.081$).

* Corresponding author. Department of Surgery, Seoul National University College of Medicine, 101 Daehak-ro, Jongno-gu, Seoul, 03080, Republic of Korea.
** Corresponding author. Division of Breast Surgery, Department of Surgery, Samsung Medical Center, Sungkyunkwan University School of Medicine, 81 Irwon-ro, Gangnam-gu, Seoul, 06351, Republic of Korea.
*** Corresponding author. Department of Surgery, Gangnam Severance Hospital, Yonsei University College of Medicine, 712 Eonjuro, Gangnam-gu, Seoul, 06273, Republic of Korea.
E-mail addresses: hblee.md@snu.ac.kr (H.-B. Lee), jaimin.ryu@samsung.com (J.M. Ryu), asg2004@yuhs.ac (S.G. Ahn).
[1] These authors contributed equally to this work.
[2] Present address: Department of Surgery, Seoul National University Hospital, Seoul National University College of Medicine, Seoul, Republic of Korea.

https://doi.org/10.1016/j.breast.2026.104738
Received 9 November 2025; Received in revised form 6 February 2026; Accepted 13 February 2026
Available online 13 February 2026

Conclusion: The machine-learning model effectively stratified patients into distinct DR risk groups and highlighted the benefit of extended ET in high-risk patients. This model supports tailored decision-making regarding extended ET in premenopausal patients with ER–positive/HER2–negative breast cancer.

1. Background

Hormone receptor-positive breast cancer accounts for approximately 60-75% of all breast cancers, making it the most common subtype. Estrogen receptor (ER)-positive/human epidermal growth factor receptor 2 (HER2)-negative breast cancer is known for having a favorable prognosis among various breast cancer subtypes [1]. However, late distant recurrence (DR), occurring more than five years after breast cancer surgery, remains a significant challenge [2]. A meta-analysis studying mortality and recurrence in patients with ER-positive breast cancer after five years of tamoxifen treatment observed a cumulative risk of delayed recurrence reaching approximately 40%, depending on stage, over 20 years after diagnosis [3].

Patients with hormone receptor-positive breast cancer typically undergo adjuvant endocrine therapy (ET) for five years after surgery, and in many cases, the treatment is extended for up to 10 years. Although extended therapy has been shown to have a positive effect on preventing recurrence, it does not impact overall survival [4,5]. However, long-term ET increases the risk of complications such as menopausal symptoms, endometrial cancer, thrombosis, and bone fractures [6]. These side effects significantly impact the patients in terms of quality of life and often serve as a crucial factor leading to treatment discontinuation [7,8]. Particularly in younger patients, treatment persistence is a critical determinant of outcomes. A recent nationwide study by Dumas et al. demonstrated that younger age is significantly associated with lower ET persistence, which directly correlates with an increased risk of recurrence [9]. Understanding these age-specific challenges in treatment adherence is essential for assessing the risk of late recurrence and optimizing long-term management in premenopausal survivors.

Therefore, it is essential to weigh the risks and benefits when considering extended ET. Predicting late recurrence informs clinical decision-making about its necessity. In postmenopausal patients, the Clinical Treatment Score at 5 years (CTS5), developed using clinical factors from the ATAC cohort and validated in the BIG 1-98 trial, has been established as a late DR risk prediction model [10]. However, several validations in premenopausal women have indicated that CTS5 has reduced predictive performance, with lower AUC values and an underestimation of risk [11–16].

Currently, clinical recommendations regarding extended ET following five years of luteinizing hormone-releasing hormone agonist-based treatment are largely extrapolated from data in postmenopausal populations. The findings from the recent EBCTCG meta-analysis, which demonstrated that extended ET significantly reduces the risk of recurrence, are also primarily based on studies of postmenopausal patients [17]. Furthermore, even within these large-scale analyses, several critical limitations have been identified by the authors: the trials were often underpowered to assess overall survival, the follow-up period after treatment completion was relatively short, and treatment adherence was not explicitly accounted for.

Given the influence of age and menopausal status on treatment outcomes in breast cancer, there is a clear need for new prediction models specifically designed for premenopausal patients. In this study, we aimed to develop a risk prediction model for late DR in premenopausal women using a machine-learning system with multicenter patients. This model could be used to evaluate the risk of late DR for individual patients with breast cancer and to determine whether extended ET would benefit high-risk patients.

2. Methods

2.1. Study population and data collection

This was a retrospective study of patients who underwent primary breast cancer surgery at multi-institutional comprehensive breast cancer centers in South Korea between January 2000 and December 2011. Patients with ER-positive/HER2-negative breast cancer aged 45 or younger were included. We excluded patients with less than 60 months of follow-up, less than 24 months of ET, or who were postmenopausal. Additionally, patients who received neoadjuvant systemic therapy, with synchronous bilateral breast cancer, or with metastatic breast cancer at diagnosis were excluded. We also excluded patients who were diagnosed with DR within five years of primary surgery. However, patients with local recurrence within five years of follow-up were included. Given that the study focused on late DR occurring between 5 and 10 years after surgery, patients with less than 60 months of follow-up were excluded. Accordingly, the effective study baseline for all analyses was established at 5 years after the primary surgery.

Postoperative surveillance was conducted in accordance with conventional clinical recommendations, with specific examinations and schedules determined by the attending physician at each institution. Generally, follow-up included annual or semi-annual physical examinations and history taking. Breast imaging, consisting of mammography and/or breast ultrasonography, was performed every 6 to 12 months. Systemic work-ups to detect distant metastasis were performed on an annual basis or as needed, depending on the patient's disease stage and the physician's clinical judgment.

In this study, we defined late DR as distant metastasis that occurred between five and 10 years after surgery. To ensure accurate machine-learning training for predicting late DR, we excluded patients who were lost to follow-up within 10 years without late DR (Table S4) or had confirmed DR after 10 years. A total of 1701 patients were finally included in the study. For the development of the extended ET model based on the late DR prediction model developed in this study, we included 1213 patients who did not receive extended ET (Fig. 1). Although the 488 patients who received extended ET were excluded from the model development cohort, they were included in the subsequent analyses to evaluate the impact of extended ET based on the predicted risk.

The data of the patients were retrospectively obtained from medical records. The pathologic staging was determined in accordance with the 8th American Joint Committee on Cancer staging manuals [18]. The necessity for obtaining informed consent was waived as the study was carried out retrospectively. The study was approved by the Institutional Review Boards of the participating institutions (IRB Nos. 2022-01-123, 3-2022-0034, and H-2112-058-1282) and conducted in accordance with the ethical principles of the Declaration of Helsinki.

2.2. Machine learning-based model development

We used a machine learning–based approach to predict late DR at 10 years by classifying patients into two groups: those who experienced DR between five and 10 years and those who did not. Four models were constructed using different combinations of predictors. The 6-feature model included age, tumor size, and number of positive lymph nodes as continuous variables, and nuclear grade, histologic grade, and progesterone receptor status as categorical variables. Two 7-feature models added either ovarian function suppression (OFS) or chemotherapy. The final 8-feature model incorporated both OFS and chemotherapy.

The patients were randomly divided into five equal-size folds, with one fold used for validation and the other four folds used for training in each iteration. Every fold was used for validation across five iterations. This 5-fold cross-validation process was repeated three times using different seeds. We reported the average values obtained from the three sets of 5-fold cross-validation.

When the variables of a patient are entered into the final 8-feature model, the output is the late DR probability value, indicating the likelihood of late DR. This late DR probability quantifies the degree of similarity between the variables of the patients and those who experienced late DR and is presented as a value between 0 and 1. A probability value approaching 1 suggests a higher similarity between the characteristics of the patient and those observed in patients who developed late DR.

2.3. Statistical analysis

Baseline characteristics were compared using the independent *t*-test for continuous variables and the chi-square or Fisher's exact test for categorical variables. Distant metastasis-free survival (DMFS) between five and 10 years was analyzed using the Kaplan-Meier method with the log-rank test. Univariable and multivariable Cox proportional hazards models were used to identify factors associated with late DR. Multicollinearity was assessed using the variance inflation factor, with a threshold of 5.0. The cut-off value of the prediction model was determined using the ROC curve and Youden's Index for binary outcomes, and the Contal and O'Quigley method for survival data [19]. All statistical analyses were performed using R version 4.2.2 (R Foundation for Statistical Computing, Vienna, Austria). Two-sided *P*-values <0.05 were considered statistically significant.

3. Results

3.1. Clinicopathologic characteristics of patients

The baseline characteristics of all 1701 patients and the extended ET model development group are summarized in Table 1. The median follow-up duration of the patients was 143.7 months (interquartile range [IQR], 128.4-167.8 months). The median age at the time of surgery was 41 years (range, 21-45). Among the 1701 patients, 108 patients (6.3%) developed late DR. Most patients had early-stage breast cancer with pathologic T1 (960 patients, 56.4%) and T2 (656 patients, 38.6%), and with pathologic N0 (993 patients, 58.4%) disease. The treatment variables included chemotherapy, extended ET, and the addition of OFS. In this study 1359 patients (79.9%) received adjuvant chemotherapy, 488 patients (28.7%) underwent extended ET, and 300 patients (17.6%) received additional OFS.

3.2. Development and performance of the late DR risk prediction model

We first developed a machine-learning model using clinicopathologic factors to predict the risk of late DR between five and 10 years after surgery. The final model, incorporating 8 clinicopathologic features, demonstrated strong predictive performance with an area under the curve (AUC) of 0.78 in the Balanced Random Forest (BRF) model. This model effectively addressed imbalanced datasets through a resampling strategy that ensured equal sample sizes between groups. Among the various machine-learning algorithms evaluated, the BRF model exhibited a stronger performance relative to the others. Detailed performance metrics for different algorithms, including logistic regression, support vector machines, and random forests, are presented in Table 2 and Fig. 2. Additionally, the performances of 6- and 7-feature models for each algorithm are summarized in Table S1.

3.3. Comparison of the risk groups stratified by the model

Using the late DR prediction model, 1701 patients were classified into 989 (58.1%) low and 712 (41.9%) high-risk groups based on the late DR probability cut-off of 0.48, as determined by methods described previously (Fig. S1). The distribution of late DR probabilities for the total patients is shown in Fig. S2.

A comparison of clinicopathologic and treatment characteristics between the low and high-risk groups revealed statistically significant differences in many variables, demonstrating that the groups were well stratified according to risk (Table S2). The median age of the low-risk group was 42 years (IQR, 39-44 years), while the high-risk group had a median age of 39 years (IQR, 36-42 years), indicating that patients in the high-risk group were significantly younger at the time of surgery. The high-risk group demonstrated significantly worse pathologic characteristics compared to the low-risk group, with a higher proportion of

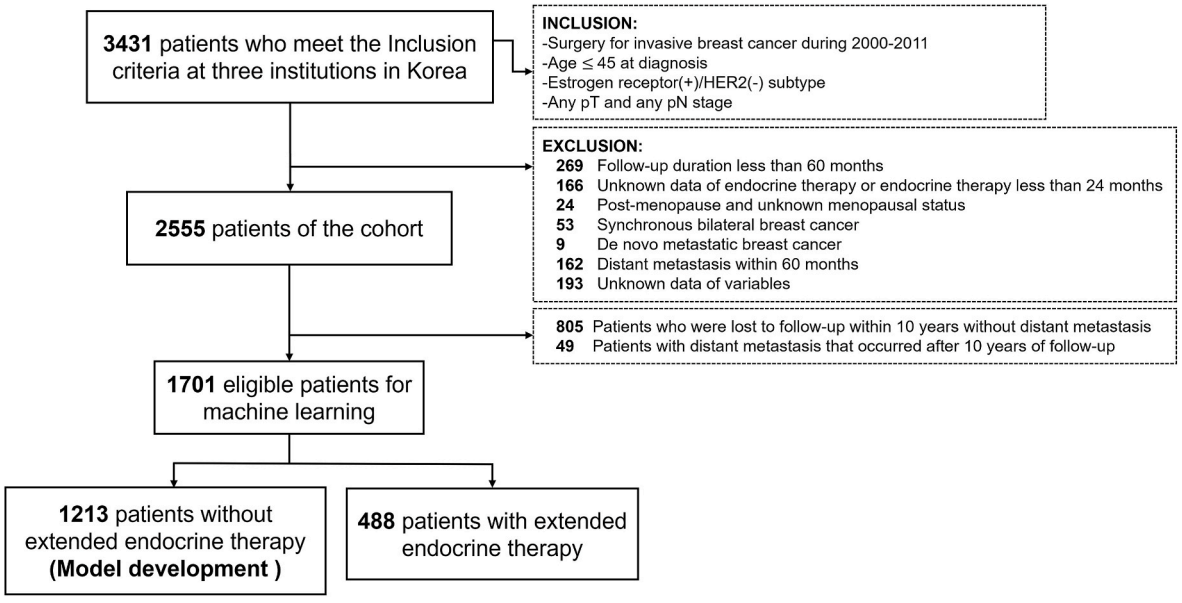

Fig. 1. Schematic diagram of the study.
[a]Patients who experienced local recurrence within five years were included in this study.

Table 1
Clinicopathologic and treatment characteristics.

Variables	Total patients (n = 1701)		Model development group (n = 1213)	
	No.	(%)	No.	(%)
Age, years				
Median	41		41	
IQR	38-43		37-43	
T stage				
pT1	960	(56.4)	752	(62.0)
pT2	656	(38.6)	408	(33.6)
pT3	83	(4.9)	51	(4.2)
pT4	2	(0.1)	2	(0.2)
N stage				
pN0	993	(58.4)	787	(64.9)
pN1	496	(29.2)	312	(25.7)
pN2	157	(9.2)	81	(6.9)
pN3	55	(3.2)	33	(2.7)
Nuclear grade				
I	250	(14.7)	204	(16.8)
II	1030	(60.5)	747	(61.6)
III	421	(24.8)	262	(21.6)
Histologic grade				
I	403	(23.7)	322	(26.5)
II	935	(55.0)	666	(54.9)
III	363	(21.3)	225	(18.5)
Progesterone receptor				
Positive	1545	(90.8)	1103	(90.9)
Negative	156	(9.2)	110	(9.1)
Adjuvant chemotherapy				
Yes	1359	(79.9)	907	(74.8)
No	342	(20.1)	306	(25.2)
Extended endocrine therapy				
Yes	488	(28.7)	0	(0.0)
No	1213	(71.3)	1213	(100.0)
Ovarian function suppression				
Yes	300	(17.6)	208	(17.1)
No	1401	(82.4)	1005	(82.9)
Radiation therapy				
Yes	1232	(72.4)	877	(72.3)
No	465	(27.3)	333	(27.5)
Unknown	4	(0.2)	3	(0.2)
Breast surgery				
BCS	1062	(62.4)	778	(64.1)
TM	639	(37.6)	435	(35.9)
Axillary surgery				
SLNB	846	(49.7)	663	(54.7)
ALND	855	(50.3)	550	(45.3)

Abbreviations: IQR, interquartile range; BCS, breast-conserving surgery; TM, total mastectomy; SLNB, sentinel lymph node biopsy; ALND, axillary lymph node dissection.

patients having advanced pathologic T and N stages (both $P < 0.001$), higher nuclear grades (nuclear grade III: 35.0% for the high-risk group vs. 17.4% for the low-risk group, $P < 0.001$) and histologic grades (histologic grade III: 32.9% for the high-risk group vs. 13.0% for the low-risk group, $P < 0.001$), and a greater percentage of progesterone receptor-negative tumors (13.2% for the high-risk group vs. 6.3% for the low-risk group, $P < 0.001$). There were also significant differences in the type of surgery performed, with a higher proportion of mastectomies (48.4% for the high-risk group vs. 31.5% for the low-risk group; $P < 0.001$) and axillary lymph node dissections (73.2% for the high-risk group vs. 37.4% for the low-risk group; $P < 0.001$) observed in the high-risk group. In addition, late DR was significantly more frequent in the high-risk group than in the low-risk group, occurring in 90 and 18 patients, respectively (12.6% vs. 1.8%; $P < 0.001$).

3.4. Cox proportional hazard regression analysis for late DR

In univariable analysis, late DR was significantly associated with younger age (HR per 1 year increment = 0.89; 95% CI, 0.86-0.92; $P < 0.001$), advanced pathologic T stages (pT2 vs. pT1: HR, 3.37; 95% CI, 2.19-5.19; $P < 0.001$; pT3 and 4 vs pT1: HR, 4.42; 95% CI, 2.22-8.83; $P < 0.001$) and N stages (pN1 vs. pN0: HR, 2.71; 95% CI, 1.72-4.26; $P < 0.001$; pN2 vs. pN0: HR, 4.87; 95% CI, 2.88-8.24; $P < 0.001$), higher histologic grade (histologic grade III vs. histologic grade I: HR, 3.23; 95% CI, 1.62-6.42; $P < 0.001$), total mastectomy (HR, 2.49; 95% CI, 1.70-3.66; $P < 0.001$, vs. breast-conserving surgery), and axillary lymph node dissection (HR, 3.61; 95% CI, 2.29-5.68; $P < 0.001$, vs sentinel lymph node biopsy). In multivariable analysis, adjusted for variables such as age, pathologic T and N stages, histologic grade, adjuvant chemotherapy, and extended ET, younger age and higher pathologic N stage were identified as significant independent factors associated with late DMFS (Table S3).

Among treatment-related variables, adjuvant chemotherapy was associated with an increased risk of late DR in univariable analysis (HR, 2.27; 95% CI, 1.22–4.24; $P = 0.010$). However, this association did not remain statistically significant in multivariable analysis (adjusted HR, 0.65; 95% CI, 0.27–1.19; $P = 0.267$). The suppression of ovarian function did not show a significant association with late DR in this study (HR, 0.87; 95% CI, 0.52–1.45; $P = 0.582$). Notably, extended ET was identified as a strong protective factor against late DR (adjusted HR, 0.31; 95% CI, 0.18–0.54; $P < 0.001$).

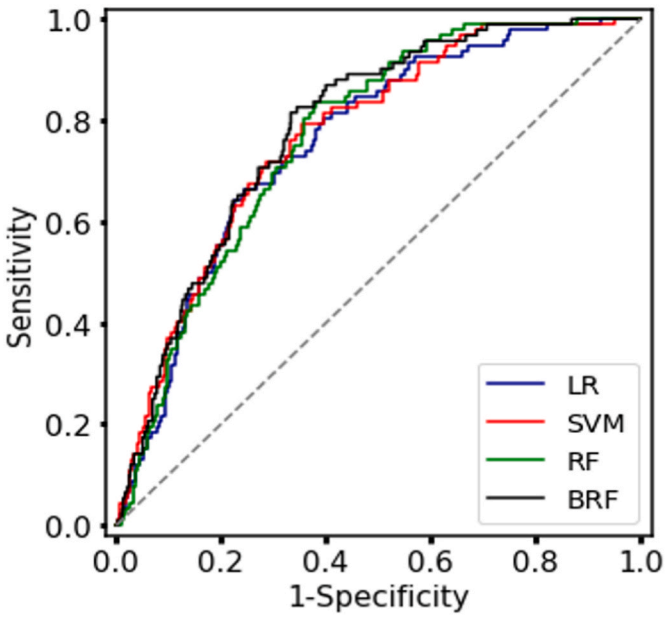

Fig. 2. ROC curves of the late distant recurrence risk prediction model.
Abbreviations: LR, logistic regression; SVM, support vector machine; RF, random forest; BRF, balanced random forest.

Table 2
Performance measures of the late distant recurrence predictive model.

Algorithms	AUC	Accuracy	Sensitivity	Specificity	Positive predict value	Negative predict value
LR	0.765	0.723	0.670	0.728	0.169	0.964
SVM	0.773	0.721	0.695	0.723	0.171	0.967
RF	0.764	0.815	0.415	0.848	0.183	0.947
BRF	0.783	0.689	0.760	0.683	0.165	0.972

Abbreviations: AUC, area under the curve; LR, logistic regression; SVM, support vector machine; RF, random forest; BRF, balanced random forest.

3.5. Late DMFS according to risks and extended ET

Kaplan-Meier survival analysis was conducted to evaluate late DMFS based on the risk stratification of the model. Patients were divided into four groups based on their late DR risk status (high-vs. low-risk) based on our late DR prediction model and whether they received extended ET. In the model development group of patients who did not receive extended ET, those classified as high-risk showed significantly worse outcomes than those classified as low-risk (Fig. 3A). When including all patients, those who were high-risk without receiving extended ET showed worse survival than those who received extended ET and all low-risk patients (Fig. 3B). Among patients in the high-risk group, extended ET was associated with a significant improvement in late DMFS compared to patients who did not receive extended ET (log-rank $P < 0.001$; HR, 0.32; 95% CI, 0.18–0.55; $P < 0.001$; Fig. 3C). However, extended ET did not result in a statistically significant improvement in late DMFS in low-risk patients (log-rank $P = 0.081$; HR, 0.20; 95% CI, 0.03–1.50; $P = 0.117$; Fig. 3D).

4. Discussion

In this large-scale retrospective multicenter study, we developed a model using a machine-learning system to predict late DR in young ER-positive/HER2-negative breast cancer patients. Our algorithm was developed to help identify high-risk groups associated with five to 10-year recurrence in young breast cancer patients and select patients who require consideration for extended ET.

Breast cancer with an ER-positive/HER2-negative subtype is known to have a relatively favorable prognosis [20,21]. However, concerns about late DR beyond five years remain [22]. With advancements in breast cancer treatment, the prognosis of other subtypes, such as HER2-positive and triple-negative breast cancer, has improved [23]. Consequently, the importance of late recurrence in ER-positive/HER2-negative breast cancer has increased. Additionally, previous reports have indicated significant differences in survival outcomes between young and old patients with luminal breast cancer [24,25]. Therefore, predicting late DR in young ER-positive/HER2-negative breast cancer patients is of utmost importance.

Nevertheless, previous methods to predict late DR in ER-positive/HER2-negative breast cancer have been validated mostly in postmenopausal women, with premenopausal women constituting only a small minority. The CTS5, which utilizes clinicopathologic features, including age, stage, and grade to predict late DR, was developed exclusively in postmenopausal women within the ATAC and BIG 1-98 cohorts [10]. In addition, the genetic profiles used in the Breast Cancer Index (BCI) and EndoPredict score, were validated with cohorts from studies such as aTTom and ABCSG6/8, where premenopausal women were largely underrepresented [26,27]. However, it has been reported that risk factors related to recurrence including pathologic, genetic, and environmental elements, vary according to menopausal status [28,29]. To address the limitations of previous studies, we constructed a multicenter cohort consisting of young patients and developed a late DR prediction algorithm.

To the best of our knowledge, our study represents the first research to develop a late DR prediction algorithm in premenopausal ER-positive/HER2-negative breast cancer patients. There is little research on the risk factors for late DR in premenopausal women. For instance, Yamashita et al. only confirmed that tumor size and lymph node metastasis are associated with late recurrence in a small cohort of 108 patients [30]. However, our study not only validated the risk factors for late DR in young patients but also developed an algorithm to predict late DR based on a large cohort of over 1700 premenopausal patients.

Another strength of our study is that our algorithm is composed solely of clinicopathologic factors that are easily obtainable in clinical practice. Tests that use multigene assays or genetic profiling are expensive and difficult to access, limiting their application in all

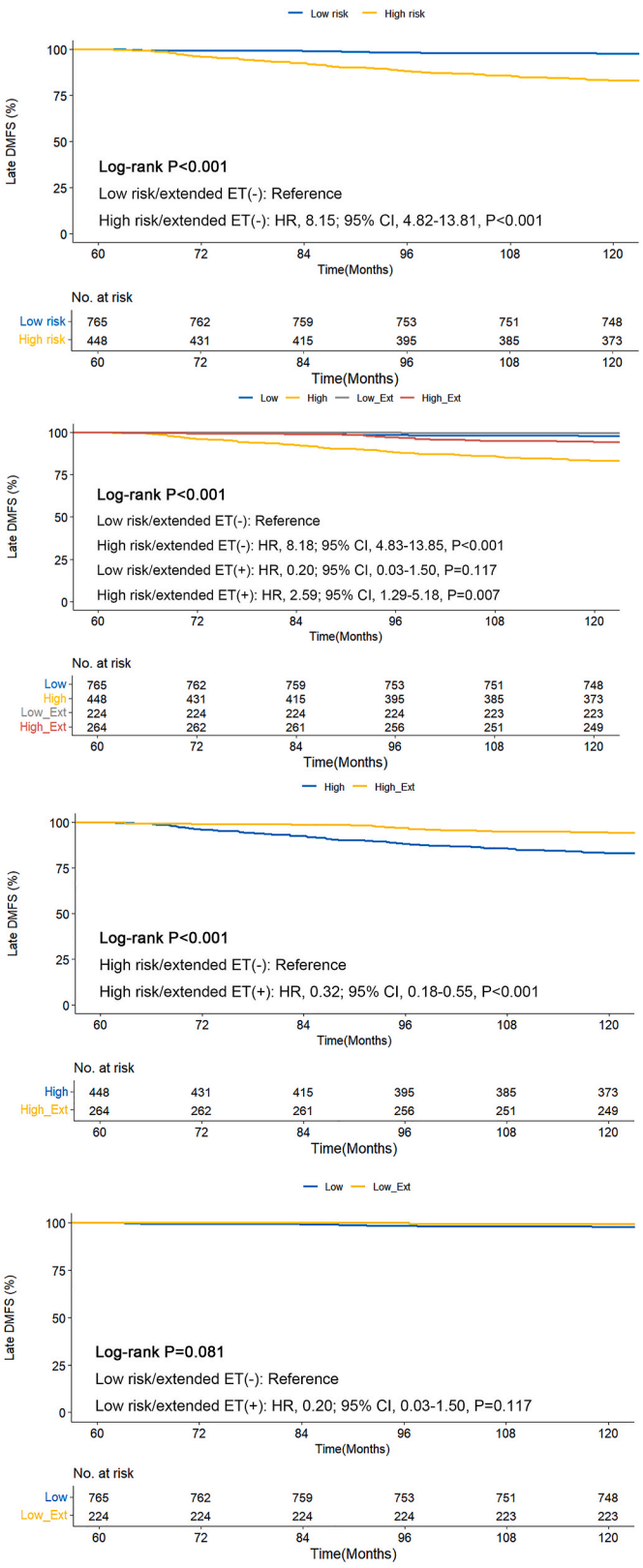

Fig. 3. Kaplan-Meier curves for late distant metastasis-free survival by stratification of the model risk groups and extended endocrine therapy. A. Model development group.
Abbreviations: ET, endocrine therapy; HR, hazard ratio; CI, confidence interval.

patients. In contrast, our model is easily accessible to anyone worldwide.

Evidence on late DR and extended ET in premenopausal ER-positive/HER2-negative breast cancer patients is limited. ATLAS and aTTom trials are the only large-scale randomized controlled trials (RCT) associated with extended ET in premenopausal women, but even in these studies, young breast cancer patients under 45 years old account for less than 20% of the enrolled population [4,5]. Therefore, our study can be considered highly significant in evaluating the role of extended ET for premenopausal patients. The validation of our algorithm in another large cohort and the implementation of an RCT will provide clearer evidence in the future.

The determination of the cut-off value for the machine-learning-based late DR prediction model developed in this study required careful consideration of multiple factors. With advancements in breast cancer treatments and improvements in survival rates, premenopausal breast cancer patients now face the prospect of a significantly longer life expectancy following diagnosis and treatment [31,32]. However, recurrences lead to additional treatment burden, psychological distress, and reduced quality of life for both patients and their families [33]. Consequently, it is critical to accurately predict the recurrence risks and develop effective prevention and follow-up strategies. Although low-risk patients have a lower overall recurrence rate, those with a late DR probability of 0.40 to 0.48 may still benefit from shared decision-making involving extended ET, given their relatively increased risk.

Multigene assays such as the BCI have been evaluated to guide decisions on extended ET. In the aTTom and IDEAL trials, extended ET significantly reduced recurrence risk in patients classified as BCI (H/I)-High, but not in those with BCI (H/I)-Low, suggesting that BCI may help identify patients who are more or less likely to benefit from extended therapy [27,34]. Integrating our late DR prediction model with multigene assays may further enhance individualized treatment strategies.

The limitation of our study is that it is based on a retrospective design, which may introduce selection bias. Nevertheless, we have built a large-scale multicenter cohort that is well-selected through the meticulous review and revision of the patient data. Additionally, our study did not include an analysis of variables such as lympho-vascular invasion and Ki-67, which could affect breast cancer prognosis. These were excluded due to unavoidable differences in interpretation criteria across institutions. Future evaluations may be possible through central laboratory reanalysis. Furthermore, because menopausal status was not reassessed at the end of the five-year period after surgery, changes in menopausal status during follow-up were not reflected in the analysis. In addition, the prediction model did not reflect potential heterogeneity related to participating centers or histologic subtypes, which should be considered a limitation of the present study. We also did not compare the performance of our machine-learning based model with multigene assays. However, given the growing importance of genomic profiling, we are collecting tissue samples for future research incorporating genomic analyses, which will provide further clinically relevant information.

5. Conclusion

In this retrospective study, we developed a late DR risk prediction model using machine-learning system for premenopausal ER-positive/HER2-negative breast cancer patients based on a large multicenter cohort. This model offers prognostic information for late DR and will assist decision-making for extended ET in each patient. An ongoing validation study of this model with a large cohort will further establish its clinical significance.

CRediT authorship contribution statement

Dong Seung Shin: Writing – review & editing, Writing – original draft, Visualization, Project administration, Methodology, Formal analysis, Data curation, Conceptualization. **Janghee Lee:** Writing – review & editing, Writing – original draft, Visualization, Project administration, Methodology, Formal analysis, Data curation, Conceptualization. **Eunhye Kang:** Writing – review & editing, Writing – original draft, Methodology, Formal analysis, Data curation, Conceptualization. **Dasom Noh:** Writing – review & editing, Writing – original draft, Methodology, Formal analysis, Data curation, Conceptualization. **Jong-Ho Cheun:** Writing – review & editing, Validation, Resources, Data curation. **Jun-Hee Lee:** Writing – review & editing, Validation, Resources, Data curation. **Yeongyeong Son:** Writing – review & editing, Validation, Resources, Data curation. **Soong June Bae:** Writing – review & editing, Validation, Resources, Data curation. **Sunyoung Kwon:** Writing – review & editing, Supervision, Project administration, Methodology, Conceptualization. **Han-Byoel Lee:** Writing – review & editing, Supervision, Project administration, Methodology, Data curation, Conceptualization. **Jai Min Ryu:** Writing – review & editing, Supervision, Project administration, Funding acquisition, Formal analysis, Conceptualization. **Sung Gwe Ahn:** Writing – review & editing, Supervision, Project administration, Funding acquisition, Conceptualization.

Data availability

The data analyzed during the current study are available from the corresponding authors upon reasonable request.

Funding

This study was supported by grant HA23C0144 from the National R&D Program for Cancer Control through the National Cancer Center funded by the Ministry of Health & Welfare, Republic of Korea (Dr Ryu) and by grant RS-2024-00343001 from the National Research Foundation of Korea (NRF), funded by the Korean government (Dr Ahn). The funders had no role in the study design, data analysis, manuscript preparation, or decision to submit for publication.

Declaration of competing interest

Han-Byoel Lee reports being a co-founder and member of the DCGen Co., Ltd board of directors, research funding from Devicor Medical Product Inc., and speaker honoraria and/or consulting fees from Alvogen, Boryung, Hologic, Intuitive, Lilly, Need, Novartis, Roche, Takeda, Celltrion Pharm, and Shin Poong Pharm, outside of the current work.

The remaining authors report no conflicts of interest.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.breast.2026.104738.

References

[1] Barzaman K, Karami J, Zarei Z, Hosseinzadeh A, Kazemi MH, Moradi-Kalbolandi S, et al. Breast cancer: biology, biomarkers, and treatments. Int Immunopharmacol 2020;84:106535. https://doi.org/10.1016/j.intimp.2020.106535.

[2] Metzger-Filho O, Sun Z, Viale G, Price KN, Crivellari D, Snyder RD, et al. Patterns of recurrence and outcome according to breast cancer subtypes in lymph node-negative disease: results from international breast cancer study group trials VIII and IX. J Clin Oncol 2013;31:3083–90. https://doi.org/10.1200/JCO.2012.46.1574.

[3] Pan H, Gray R, Braybrooke J, Davies C, Taylor C, McGale P, et al. 20-Year risks of breast-cancer recurrence after stopping endocrine therapy at 5 years. N Engl J Med 2017;377:1836–46. https://doi.org/10.1056/NEJMoa1701830.

[4] Davies C, Pan H, Godwin J, Gray R, Arriagada R, Raina V, et al. Long-term effects of continuing adjuvant tamoxifen to 10 years versus stopping at 5 years after diagnosis of oestrogen receptor-positive breast cancer: ATLAS, a randomised trial. Lancet 2013;381:805–16. https://doi.org/10.1016/S0140-6736(12)61963-1.

[5] Gray RG, Rea D, Handley K, Bowden SJ, Perry P, Earl HM, et al. aTTom: long-term effects of continuing adjuvant tamoxifen to 10 years versus stopping at 5 years in 6,953 women with early breast cancer. J Clin Oncol 2013;31:5. https://doi.org/10.1200/jco.2013.31.18_suppl.5.

[6] Francis PA, Pagani O, Fleming GF, Walley BA, Colleoni M, Lang I, et al. Tailoring adjuvant endocrine therapy for premenopausal breast cancer. N Engl J Med 2018; 379:122–37. https://doi.org/10.1056/NEJMoa1803164.
[7] Bernhard J, Luo W, Ribi K, Colleoni M, Burstein HJ, Tondini C, et al. Patient-reported outcomes with adjuvant exemestane versus tamoxifen in premenopausal women with early breast cancer undergoing ovarian suppression (TEXT and SOFT): a combined analysis of two phase 3 randomised trials. Lancet Oncol 2015;16: 848–58. https://doi.org/10.1016/S1470-2045(15)00049-2.
[8] Ganz PA, Cecchini RS, Julian TB, Margolese RG, Costantino JP, Vallow LA, et al. Patient-reported outcomes with anastrozole versus tamoxifen for postmenopausal patients with ductal carcinoma in situ treated with lumpectomy plus radiotherapy (NSABP B-35): a randomised, double-blind, phase 3 clinical trial. Lancet 2016;387: 857–65. https://doi.org/10.1016/S0140-6736(15)01169-1.
[9] Dumas E, Jochum F, Coussy F, Hamy AS, Majdling A, Houzard S, et al. Explaining the relationships between age, endocrine therapy persistence, and risk of recurrence in hormone receptor-positive early breast cancer: a nationwide cohort study. J Clin Oncol 2025;43:1863–74. https://doi.org/10.1200/jco.24.01131.
[10] Dowsett M, Sestak I, Regan MM, Dodson A, Viale G, Thurlimann B, et al. 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;36:1941–8. https://doi.org/10.1200/JCO.2017.76.4258.
[11] Lee JH, Lee SK, Chae BJ, Yu J, Lee JE, Kim SW, et al. Validation of the clinical treatment score post-five years in breast cancer patients for predicting late distant recurrence: a single-center investigation in Korea. Front Oncol 2021;11:691277. https://doi.org/10.3389/fonc.2021.691277.
[12] Dejthevaporn T, Patanayindee P. Clinical treatment score Post-5 years as a tool for risk estimation of late recurrence in Thai patients with estrogen-receptor-positive, early breast cancer: a validation study. Breast Cancer 2023;17: 11782234231186869. https://doi.org/10.1177/11782234231186869.
[13] Lee J, Cha C, Ahn SG, Kim D, Park S, Bae SJ, et al. Validation of clinical treatment score post-5 years (CTS5) risk stratification in premenopausal breast cancer patients and Ki-67 labelling index. Sci Rep 2020;10:16850. https://doi.org/10.1038/s41598-020-74055-3.
[14] Villasco A, Agnelli F, D'Alonzo M, Accomasso F, Sismondi P, Biglia N. Validation of CTS5 on a retrospective cohort of real-life Pre- and postmenopausal patients diagnosed with estrogen receptor-positive breast cancers: is it prognostic? Clin Breast Cancer 2021;21:e53–62. https://doi.org/10.1016/j.clbc.2020.06.008.
[15] Ning L, Liu Y, He X, Han R, Xin Y, Zhao J, et al. Validation of CTS5 model in large-scale breast cancer population and combination of CTS5 and Ki-67 status to develop a novel nomogram for prognosis prediction. Am J Clin Oncol 2024;47: 228–38. https://doi.org/10.1097/COC.0000000000001080.
[16] Richman J, Ring A, Dowsett M, Sestak I. Clinical validity of clinical treatment score 5 (CTS5) for estimating risk of late recurrence in unselected, non-trial patients with early oestrogen receptor-positive breast cancer. Breast Cancer Res Treat 2021;186: 115–23. https://doi.org/10.1007/s10549-020-06013-6.
[17] Extending the duration of endocrine treatment for early breast cancer: patient-level meta-analysis of 12 randomised trials of aromatase inhibitors in 22 031 postmenopausal women already treated with at least 5 years of endocrine therapy. Lancet 2025;406:603–14. https://doi.org/10.1016/s0140-6736(25)01013-x.
[18] Amin MB, Edge SB, Greene FL, Byrd DR, Brookland RK, Washington MK, et al. AJCC cancer staging manual. eighth ed. Springer; 2017.
[19] Contal C, O'Quigley J. An application of change point methods in studying the effect of age on survival in breast cancer. Comput Stat Data Anal 1999;30:253–70. https://doi.org/10.1016/S0167-9473(98)00096-6.
[20] Hwang KT, Kim J, Jung J, Chang JH, Chai YJ, Oh SW, et al. Impact of breast cancer subtypes on prognosis of women with operable invasive breast cancer: a population-based study using SEER database. Clin Cancer Res 2019;25:1970–9. https://doi.org/10.1158/1078-0432.Ccr-18-2782.
[21] Onitilo AA, Engel JM, Greenlee RT, Mukesh BN. Breast cancer subtypes based on ER/PR and Her2 expression: comparison of clinicopathologic features and survival. Clin Med Res 2009;7:4–13. https://doi.org/10.3121/cmr.2009.825.
[22] Jatoi I, Anderson WF, Jeong JH, Redmond CK. Breast cancer adjuvant therapy: time to consider its time-dependent effects. J Clin Oncol 2011;29:2301–4. https://doi.org/10.1200/jco.2010.32.3550.
[23] Cossetti RJ, Tyldesley SK, Speers CH, Zheng Y, Gelmon KA. Comparison of breast cancer recurrence and outcome patterns between patients treated from 1986 to 1992 and from 2004 to 2008. J Clin Oncol 2015;33:65–73. https://doi.org/10.1200/jco.2014.57.2461.
[24] Johansson ALV, Trewin CB, Hjerkind KV, Ellingjord-Dale M, Johannesen TB, Ursin G. Breast cancer-specific survival by clinical subtype after 7 years follow-up of young and elderly women in a nationwide cohort. Int J Cancer 2019;144: 1251–61. https://doi.org/10.1002/ijc.31950.
[25] Partridge AH, Hughes ME, Warner ET, Ottesen RA, Wong YN, Edge SB, et al. Subtype-dependent relationship between young age at diagnosis and breast cancer survival. J Clin Oncol 2016;34:3308–14. https://doi.org/10.1200/jco.2015.65.8013.
[26] Dubsky P, Brase JC, Jakesz R, Rudas M, Singer CF, Greil R, et al. The EndoPredict score provides prognostic information on late distant metastases in ER+/HER2-breast cancer patients. Br J Cancer 2013;109:2959–64. https://doi.org/10.1038/bjc.2013.671.
[27] Bartlett JMS, Sgroi DC, Treuner K, Zhang Y, Ahmed I, Piper T, et al. Breast cancer index and prediction of benefit from extended endocrine therapy in breast cancer patients treated in the adjuvant Tamoxifen-To offer more? (aTTom) trial. Ann Oncol 2019;30:1776–83. https://doi.org/10.1093/annonc/mdz289.
[28] Yoshimoto N, Nishiyama T, Toyama T, Takahashi S, Shiraki N, Sugiura H, et al. Genetic and environmental predictors, endogenous hormones and growth factors, and risk of estrogen receptor-positive breast cancer in Japanese women. Cancer Sci 2011;102:2065–72. https://doi.org/10.1111/j.1349-7006.2011.02047.x.
[29] Yamamoto Y, Yamamoto-Ibusuki M, Iwase H. Menopausal status should be taken into consideration for patients with luminal a breast cancer in terms of the effect of differential biology on prognosis. J Clin Oncol 2013;31:2516. https://doi.org/10.1200/jco.2013.49.4062.
[30] Yamashita H, Ogiya A, Shien T, Horimoto Y, Masuda N, Inao T, et al. Clinicopathological factors predicting early and late distant recurrence in estrogen receptor-positive, HER2-negative breast cancer. Breast Cancer 2016;23:830–43. https://doi.org/10.1007/s12282-015-0649-0.
[31] Lee SB, Ko SS, Park CH, Nam SJ, Choi JE, Jung YS, et al. Chronologically changing patterns in the survival of Korean patients with breast cancer and related clinical factors: a nationwide registry-based study. Breast Cancer Res Treat 2018;172: 273–82. https://doi.org/10.1007/s10549-018-4892-x.
[32] Kim H, Lee SB, Kim J, Chung IY, Kim HJ, Ko BS, et al. Improvement of survival in Korean breast cancer patients over a 14-year period: a large-scale single-center study. PLoS One 2022;17:e0265533. https://doi.org/10.1371/journal.pone.0265533.
[33] Northouse LL, Mood D, Kershaw T, Schafenacker A, Mellon S, Walker J, et al. Quality of life of women with recurrent breast cancer and their family members. J Clin Oncol 2002;20:4050–64. https://doi.org/10.1200/jco.2002.02.054.
[34] Noordhoek I, Treuner K, Putter H, Zhang Y, Wong J, Meershoek-Klein Kranenbarg E, et al. Breast cancer index predicts extended endocrine benefit to individualize selection of patients with HR(+) early-stage breast cancer for 10 years of endocrine therapy. Clin Cancer Res 2021;27:311–9. https://doi.org/10.1158/1078-0432.Ccr-20-2737.