January 2015

Development And Validation Of A Nomogram To Predict Pathological Complete Response To Neoadjuvant Chemotherapy For Breast Cancer

Qi Yang
Yale University, qi.yang@yale.edu

Development and Validation of a Nomogram to Predict Pathological Complete Response to Neoadjuvant Chemotherapy for Breast Cancer

Victoria Qi Yang

Abstract

Purpose The purpose of this study was to combine clinical pathologic variables that are associated with pathological complete response (pCR) following neoadjuvant chemotherapy into a prediction nomogram.

Methods A total of 15,553 women who underwent neoadjuvant chemotherapy for invasive breast cancer in 2010 and 2011 were identified from National Cancer Database (NCDB). Univariate analysis and multivariate logistic regression analysis were used to examine the association of patient age, race, tumor histology, tumor grade, molecular type, and clinical stage with pCR. A nomogram was then developed to predict individual patient probability of pCR to neoadjuvant chemotherapy. Internal validation was performed in terms of discrimination and calibration. The nomogram was then tested against 319 patients from Yale New-Haven Hospital.

Results The predicted probability of the nomogram is between 4% and 74% based on clinical characteristics. In multivariate analysis, high pCR rate is significantly associated with young age, white race, ductal carcinoma, poorly differentiated tumor, Her2 positive and triple negative tumor, small tumor size and less advanced nodal status ($p<0.001$). The nomogram had the area under the curve (AUC) of 0.697 in the training set, 0.693 in the internal validation set, and 0.798 in the external validation set. The calibration plot showed good agreement between predicted and actual outcomes.

Conclusions We developed a nomogram that can be used to predict the individual probability of achieving pCR following neoadjuvant chemotherapy in patients with invasive breast cancer, based upon age, race and clinicopathologic characteristics.

Introduction

Breast cancer is the most common cancer affecting women and the second most common cause of cancer-related mortality [1]. Neoadjuvant chemotherapy is a treatment given before surgery to patients who have high-risk early-stage breast cancer. In recent years, neoadjuvant chemotherapy is considered as the standard of care for locally advanced and inoperable tumors [2], and it is increasingly used in the management of early stage breast cancer because of a number of potential benefits. First, neoadjuvant chemotherapy may shrink a breast tumor from its current inoperable state to a smaller size, which may allow subsequent surgery to remove the tumor [3]. Second, neoadjuvant chemotherapy permits breast-conserving surgery and a better cosmetic outcome in patients who otherwise would need a mastectomy [4, 5, 31]. Third, neoadjuvant chemotherapy provides a real-time evaluation and early observation of tumor response to treatment, which may lead to modifications of the treatment plan or discontinuation of ineffective therapy in the event of poor response. Fourth, neoadjuvant chemotherapy may provide prognostic information and allow investigators an opportunity to examine modulation of tissue biomarkers and imaging from the time of biopsy to the time of definitive breast surgery [6]. Pathological complete response (pCR) is considered as a valid early surrogate of long-term outcome and cure from breast cancer and has been used as an endpoint in trials of new types of chemotherapy for breast cancer [6, 7, 8].

Histological type, grade, tumor size, lymph node involvement, estrogen receptor (ER) and Her2 status all influence prognosis and the probability of response to systemic treatment. These clinical and pathological factors have correlate with the recurrence rate and prognosis of breast cancer for patients with and without adjuvant chemotherapy. In recent years, general agreement has been reached that these factors also correlate with the rate of pCR following

neoadjuvant chemotherapy. High pCR rate has been observed among patients with ER negativity, high tumor grade, high tumor proliferative activity, and small tumor size. Two large clinical trials sponsored by NCI reported that by integrating molecular diagnostic information into clinical decision-making, patients and clinicians will be able to make more informed decision regarding the most appropriate treatment options and benefit from chemothearpy.[9, 10]. However, most of the trials have been small and none include all the possible molecular markers (i.e. estrogen receptor, progesterone receptor, Her2) of patients with breast cancer into one model to estimate the probability of pCR to neoadjuvant chemotherapy.

To date, there are only three nomograms developed to predict probability of achieving pCR to neoadjuvant chemotherapy for breast cancer [2, 11, 12]. All of them have demonstrated high accuracy. Unfortunately, small sample size (<600 patients) and single data source limit the generalizability of those studies. Furthermore, previous studies only included patients who completed three or four courses of chemotherapy; patients who started receiving neoadjuvant chemotherapy but were not able to complete a full course were excluded. The purpose of this study was to combine clinical pathologic variables that are associated with pCR following neoadjuvant chemotherapy into a prediction nomogram. Our nomogram is strengthened by the large sample size and the inclusion of patient's ER, progesterone receptor (PR), and Her 2 status. Also, our study included patients who received neoadjuvant chemotherapy regardless of chemotherapy completion, as in practice completion would not be known in advance. Therefore, we consider this nomogram a robust and accurate tool to estimate the probability of benefit from neoadjuvant chemotherapy for an individual patient in the real world.

Methods

Study population

The National Cancer Database (NCDB) is a joint project of the American Cancer Society and the Commission on Cancer of the American College of Surgeons. It is a national oncology outcomes database for more than 1500 Commission-accredited cancer programs in the United States. It collects data annually from a broad range of hospitals throughout the United States on a voluntary basis. About 70 percent of the newly diagnosed cases of cancer in the United States are captured at the institutional level and reported to the NCDB through a computerized format using coding schema from the *Data Acquisition Manual* [13], the *American Joint Commission on Cancer (AJCC) Cancer Staging* Manual [14], and the World Health Organization's *International Classification of Disease for Oncology* (ICD-O-2/3) system for coding site and histologic type [15]. An algorithm based on patient and disease characteristics, including patent gender, site, date of birth, and zip code, was used to identify and remove duplicate records to ensure that patients seen at multiple institutions for the same cancer were not included within the database more than once [16].

Cases to be included in this study were extracted from the 334,447 females with invasive breast cancer in the NCDB diagnosed in 2010 and 2011. In total 29,534 women underwent neoadjuvant chemotherapy for invasive breast cancer in 2010 and 2011 but many had unknown pathological response. Among 19,310 women where the pathological response was known, 6,244 (32%) had complete response (pCR), 11,522 (60%) had partial response, and 1,544 (8%) had no response. 15,553 women had known data for all the seven predictors. The NCDB variable indicating the sequence of systemic therapy and surgery was used to determine the timing of chemotherapy. Patients who received only neoadjuvant hormone therapy or neoadjuvant radiation therapy

were excluded, but patients who received both neoadjuvant chemotherapy and neoadjuvant hormone therapy or neoadjuvant radiation therapy were not excluded from the study sample, and they represent 6.3% and 1.4% of the total, respectively. Eighty percent of the data (12,442 patients; training set) was then randomly selected and used to develop the nomogram, and the remaining 20% (3,111 patients; validation set) was used for internal validation. 319 patients diagnosed with invasive breast cancer between 2006 and 2012 was then analyzed for external validation. The sample selection procedure is shown in **Figure 1**.

Histologies were classified according to ICD-O-3 codes. The three main histologic groupings, representing invasive breast cancer, were: invasive ductal carcinoma, invasive lobular carcinoma, and mixed ductal and lobular carcinoma. The remaining types of histology were categorized as “other”.

The ICD-O-3 grading system was used with four separate categories: well differentiated (most like normal tissue), moderately well differentiated, poorly differentiated, and undifferentiated (least like normal tissue). In this study, poorly differentiated and undifferentiated tumor were classified in the poorly differentiated group. The grade information was from the final pathologic diagnosis, otherwise from the microscopic description or comments if the differentiation was not available in the final pathologic diagnosis.

AJCC staging, which is designated by tumor, node and metastasis classification, was used to describe the extent of disease. Since our study involved neoadjuvant chemotherapy, we used the clinical stage which was estimated prior to chemotherapy treatment rather than the pathological stage obtained at surgery.

In this study, breast cancer was classified into four main molecular subtypes based on immunohistochemistry ER/PR and Her2 expression, positive and/or negative. The four groups

are HR+/Her2+, HR+/Her2−, HR−/ Her2+, and HR−/Her2− [17]. HR+ refers to either ER positive or PR positive, HR- refers to both ER and PR negative. Each group has a distinct prognosis, and unique molecular portrait that governs tumor progression [18, 19, 20].

In addition to histology, tumor grade, molecular type, and staging, data regarding patient characteristics including age, race, income, insurance status, facility type and location were also collected. Age was recorded at the patient's last birthday before diagnosis. Race was grouped into White, Black, and other (e.g. American Indian, Asian, Pacific Islander). Patient income was based on the median family income of the patient's zip code of residency at the time of diagnosis as per the US Census. Income was categorized as quartiles based on equally proportioned income ranges among all US zip codes. Patient's primary insurance carrier was identified at the time of initial diagnosis and /or treatment. Facility location were grouped into nice geographic regions: New England, Middle Atlantic, South Atlantic, East North Central, East South Central, West North Central, West South Central, Mountain, and Pacific. Patient characteristics are shown in **Table 1**.

Statistical analyses

Descriptive statistics focused on frequencies and proportions for all the categorical variables. Univariate analysis with chi-square tests and multivariate logistic regression analysis were used to test the association of predictors to pCR following neoadjuvant chemotherapy. Variables which are clinically relevant, including age, histology, tumor grade, molecular type, clinical T stage and clinical N stage were selected and included in the multivariate logistic model. Race was also included in the multivariate analyses because our prior work has demonstrated its association with pCR [31]. Odds ratios were calculated for each independent variable. The

regression coefficients from the multivariate logistic regression model were then used to construct the nomogram that predicts the probability of achieving pCR for an individual patient. We also tested Interactions between covariates.

The area under the receiver operating characteristic curve (ROC) was used to quantify the nomogram's predictive accuracy. An ROC curve plots the true positive fraction (sensitivity) against the false positive fraction (1-specificity) at different predicted risk thresholds. ROC curves were constructed for both the training set (contains 80% of the data) and validation set (contains the remaining 20% of the data), respectively. The AUC value is between 0.5 and 1 [21]. AUC=1 means perfect accuracy because both the sensitivity and specificity are 1 so there are no false positives and no false negatives [22]. AUC=0.5 means the test discriminates patients who achieved pCR and patients who did not achieve pCR by chance. Since the sample size is large, bootstrapping method was not performed/needed.

Model calibration was also constructed to study the relationship between the actual probabilities and the predicted probabilities. The predicted probability provided by the nomogram for each patient was ranked and grouped into ten quantiles. The mean predicted probability was then calculated for each quantile and compared with the actual probability. The perfect calibration curve is that all the data points laid on the regression line $y=x$ ($\alpha=0$ and $\beta=1$), in which predicted and actual probabilities are identical. All analyses were carried out in SAS Version 9.3. All tests were two sided, with a significance level of 0.05.

Results

Out of 334,447 cases of invasive breast cancer diagnosed in 2010 and 2011, 29,534 (8.8%) underwent neoadjuvant chemotherapy, including 2,052 (6.9%) who also received neoadjuvant hormonal therapy and 616 (2.1%) who also received neoadjuvant radiation therapy. Among patients with known pathological response, 15,553 patients had known data for all seven predictors.

Out of 15,553 patients who underwent neoadjuvant chemotherapy and had non-missing data for all the covariates, 4915 patients (32%) achieved pCR. The relationship between pCR and baseline patient/clinicopathological characteristics are shown in **Table 1**. The mean age of patients was 52.2 years (median, 52.0 years), with a range of 18 to 90 years. 91% of patients were between age 30 and 69, the age groups at highest risk of being diagnosed with breast cancer. As expected, there is a slightly higher incidence of white women (77%) than in the US population (74.8% in the 2010 census [http://www.census.gov]).

On univariate analysis, women who had pCR tended to be younger and with higher tumor grade. Patients with ductal carcinoma (34%) were more likely to have pCR compared to those with lobular carcinoma (14%) or mixed ductal and lobular carcinoma (18%). Patients who had molecular type of HR-/Her2+ achieved pCR 50% of the time, while only 18% of patients with molecular type of HR+/Her2- achieved pCR. Interestingly, 40% of patients achieved pCR with triple negative breast cancer. Achieving pCR was clearly associated with smaller tumor size: 41% of patients with clinical T1 tumor versus 35% for T2, and 25% for T3 (except T4 tumors since they are tumors that are invading the skin or chest wall and having multiple satellite nodules, therefore are not classified based on size of the primary tumor). The same trend was observed in clinical N stage, where achieving pCR was associated with less advanced nodal disease. There

is no significant difference observed among white, black and other races in univariate analysis. Forty-four percent of the patients resided in areas with median household income over $46,000. The majority of patients had private insurance (64%) and received treatment from a comprehensive community cancer program (56%).

Eighty percent, or 12,442 patients, were included in the training set, which was used to construct the nomogram. On multivariate analysis, all seven predictors were independently associated with pCR. We excluded insurance status and included race in the multivariate regression model based on clinical considerations and ethical concerns. (**Table 2**). Young age groups (<30, 30-39, 40-49, and 50-59 years) showed better outcome and reached significance in the multivariate model (adjusted OR: 2.11, 2.05, 1.92, and 1.69, respectively). White women were more likely to achieve pCR than black women (adjusted OR: 1.18, 95% CI: 1.06-1.32). In regard to tumor histology, patients with ductal carcinoma had significantly higher pCR rate compared with patients with lobular carcinoma (adjusted OR: 1.63, 95% CI: 1.27-2.09). Patients with poorly differentiated tumor were 1.78 times more likely to achieve pCR than those with well differentiated tumor. Patients with triple negative breast cancer were 2.37 times more likely to achieve pCR than patients with HR+/Her2- cancer, which increased to a 2.60-fold higher rate in patients with HR+/Her2- cancer and a 4.00-fold higher rate in patients with HR-/Her2+ cancer after adjusting for other variables. Patients with smaller tumor size and less advanced nodal status were significantly more likely to achieve pCR. Compared with patients who had T3 tumor, there was an increased pCR rate observed in patients with T2 tumor (adjusted OR: 1.47, 95% CI: 1.32-1.63) and patients with T1 tumor (adjusted OR: 2.08, 95% CI: 1.81-2.38). Patients with clinical N0 disease is associated with a 45% increase in the odds of pCR compared to those who had clinical N3 disease.

A nomogram to predict probability of pCR to neoadjuvant chemotherapy based on patient age, race, tumor histology, tumor grade, molecular type, and clinical stage is shown in **Figure 2**. The predicted probability is between 4% and 74%. Probability less than 10% is considered less likely to achieve pCR and may not benefit from neoadjuvant chemotherapy, and over 30% is considered more likely to achieve pCR and benefit from neoadjuvant chemotherapy. Molecular type has the largest impact on the probability of pCR and therefore 100 points was assigned. The nomogram's predictive accuracy was quantified by ROC curve (**Figure 3**). The AUC is 0.697 for the training set (n=12,442). Internal validation was performed by using the remaining 20% of the data (n=3,111) and the AUC is 0.693 for the validation set. External validation of the model using 319 cases treated at Yale Cancer Center produced an AUC of 0.798. We next examined the relationship between the nomogram-predicted probability of pCR to the actual probability in both training and validation sets (**Figure 4**). The slope of the linear regression line for the training set is 1.00 (R^2=0.9926), 0.95 for the internal validation set (R^2=0.9795), and 1.33 for the external validation set (R^2=0.9982), which indicated an accurate prediction of pCR. These results demonstrate that the nomogram was well calibrated to predict the probability of achieving pCR for individual patients by integrating breast cancer molecular types with other routinely available variables.

Discussion

In the present study, we developed a nomogram to predict the probability of achieving pCR in breast cancer patients receiving neoadjuvant chemotherapy. This predictive and prognostic model is internally validated and showed good performance in terms of discrimination and

calibration. This user friendly nomogram would be useful for risk assessment and could be the basis for individualized risk-adaptive therapy.

To date, pCR has been proposed as a surrogate endpoint for prediction of long-term survival and cure from breast cancer [23] Several large randomized studies have shown that patients achieving pCR to chemotherapy have better long-term survival than those who respond incompletely to primary chemotherapy [7, 8, 24]. Four nomograms have been reported in breast cancer patients receiving neoadjuvant chemotherapy [2, 11, 12, 25], and three of them were to predict the probability of pCR with neoadjuvant chemotherapy [2, 11, 12]. The clinicopathological factors that have been considered in previous studies include histologic grade, ER status, Ki-67, clinical stage, pathologic stage, and number of chemotherapy cycles. Differing from the previous nomograms, we combined estrogen receptor, progesterone receptor and Her2 statuses and integrated the newly created variable, molecular type, into our predictive model. In our results, patients with Her2 positive breast cancer were significantly more likely to achieve pCR than those with Her2 negative.

Several limitations may be considered when interpreting our results. First, the variables in the NCDB dataset do not allow us to distinguish the type of chemotherapy. Further research including type of chemotherapy would achieve better predictions. For example, NeoALLTO investigators previously reported that the combination of paclitaxel, lapatinib, and trastuzumab significantly increased the pCR rate compared to paclitaxel combined with either drug alone [26, 27]. Secondly, the study sample was representative of the US breast cancer population, dominated by whites and patients with relatively high income. A study reported that epidemiology and tumor biology of the Asian breast cancer patients is somewhat different from those of the Westerners [28, 29]. Although the pCR rate was nearly the same among white, black and other races in univariate analysis, the multivariate result might be biased away from

the null hypothesis. This is because blacks get breast cancer at a younger age and have more high grade and triple negative cancers. Therefore, we would expect them to have a higher pCR rate than whites.

These limitations are balanced by several study strengths including large sample size, enrollment of patients from diverse facilities, and inclusion of molecular type in the predictive model. To our knowledge, this is the first study looking at the relationship between neoadjuvant chemotherapy and the individual probability of achieving pCR on the national level and is the largest series reporting pCR outcome of neoadjuvant chemotherapy for breast cancer. We found that high pCR rate is associated with young age, white race, ductal carcinoma, high tumor grade, Her2 positive and triple negative tumors, smaller tumor size, and less advanced nodal status. This is consistent with previous findings from smaller single institution studies. Our large sample size, including 15,553 patients, allowed us to use approximated regression line directly instead of bootstrap sampling method, therefore avoided the built-in errors. Most importantly, this is the first study to integrate estrogen receptor, progesterone receptor and Her2 statuses as one variable into one predictive model. We found that poorly differentiated and Her2 positive tumors are more chemosensitive and are more likely to associate with higher probability of pCR regardless of ER/PR status. On the contrary, well-differentiated and Her2 negative tumors are less likely to achieve pCR after neoadjuvant chemotherapy. Fourth, our study included patients treated with different number of neoadjuvant chemotherapy cycles, not only three or four courses. Several studies demonstrated that patients with four courses of neoadjuvant anthracycline-based combination chemotherapy were more likely to achieve pCR than those with three courses [2, 11]. In general, patients who can complete four courses are more likely to have better outcome than those who received less courses.

In addition, although pCR is considered as a valid early surrogate of long-term survival from breast cancer and studies have found that Her2 positive tumor is significantly associated with high pCR rate, a retrospective analysis which included 1,731 patients with noninflammatory breast cancer demonstrated that progression-free survival rates were significantly worse for Her2 positive breast cancer in both hormone receptor positive and negative groups [30]. Therefore, the relationship between pCR rate, Her2 status and long-term survival warrants further investigation.

It is also believed that higher probability of pCR might associate with higher rate of breast conservation. A recent study observed a strong positive correlation between pCR and lumpectomy rate in patients with aggressive breast cancer subtypes, including Her2 positive tumors and triple negative tumors [31]. Currently, the type of surgery is chosen mostly according to baseline tumor characteristics prior to neoadjuvant therapy [26]. Several international expert panels have recommended that the rate of breast conservation surgery should increase in patients who respond well to neoadjuvant chemotherapy. A nomogram was also developed to predict the probability of successful conservative surgery with neoadjuvant chemotherapy [25].

In summary, we developed a nomogram that can be used to predict the individual probability of achieving pCR following neoadjuvant chemotherapy in patients with invasive breast cancer, based upon age, race and clinicopathologic characteristics. The nomogram may be useful to aid clinicians to make individualized treatment plans for patients by estimating the potential benefit from neoadjuvant chemotherapy. The emerging field of molecular marker research may substantially improve nomogram predictions. In the future, expectations of more accurate and specific nomograms may be justified.

Table 1. Patient Characteristics according to Pathological Complete Response

Patient Characteristics	Total	pCR+	pCR-	p-value
	15,553	4,915 (32%)	10,638 (68%)	
Demographic Factors				
Age				<0.0001
<30	273 (2%)	103 (38%)	170 (62%)	
30-39	2,031 (13%)	729 (36%)	1,302 (64%)	
40-49	4,329 (28%)	1,452 (34%)	2,877 (66%)	
50-59	4,706 (30%)	1,510 (32%)	3,196 (68%)	
60-69	3,082 (20%)	841 (27%)	2,241 (73%)	
70-79	946 (6%)	237 (25%)	709 (75%)	
80+	186 (1%)	43 (23%)	143 (77%)	
Race				0.3038
White	12,022 (77%)	3,814 (32%)	8,208 (68%)	
Black	2,696 (17%)	824 (31%)	1,872 (69%)	
Other	835 (5%)	277 (33%)	558 (67%)	
Income				0.0071
<$30,000	1,820 (12%)	565 (31%)	1,255 (69%)	
$30,000-$34,999	2,368 (16%)	678 (29%)	1,690 (71%)	
$35,000-$45,999	4,040 (28%)	1,309 (32%)	2,731 (68%)	
$46,000+	6,462 (44%)	2,079 (32%)	4,383 (68%)	
Insurance				<0.0001
None	640 (4%)	194 (30%)	446 (70%)	
Private	10,002 (64%)	3,357 (34%)	6,645 (66%)	
Medicaid	2,066 (13%)	579 (28%)	1,487 (72%)	
Medicare	2,484 (16%)	665 (27%)	1,819 (73%)	
Other Government	185 (1%)	62 (34%)	123 (66%)	
Unknown	176 (1%)	58 (33%)	118 (67%)	
Location				0.0795
New England	867 (6%)	263 (30%)	604 (70%)	
Middle Atlantic	2,041 (13%)	589 (29%)	1,452 (71%)	
South Atlantic	3,643 (23%)	1,183 (32%)	2,460 (68%)	
East North Central	2,517 (16%)	824 (33%)	1,693 (67%)	
East South Central	851 (5%)	254 (30%)	597 (70%)	
West North Central	1,298 (8%)	410 (32%)	888 (68%)	
West South Central	1,366 (9%)	426 (31%)	940 (69%)	
Mountain	770 (5%)	239 (31%)	531 (69%)	
Pacific	2,200 (14%)	727 (33%)	1,473 (67%)	
Facility Type				0.0072
Community Cancer Program	1,256 (8%)	344 (27%)	912 (73%)	
Comprehensive Community	8,760 (56%)	2,781 (32%)	5,979 (68%)	
Academic/Research Program	5,484 (35%)	1,771 (32%)	3,713 (68%)	
Other Specified Types of Cancer Programs	53 (0%)	19 (36%)	34 (64%)	

Year of Diagnosis				0.0865
2010	7,679 (49%)	2,377 (31%)	5,302 (69%)	
2011	7,874 (51%)	2,538 (32%)	5,336 (68%)	
Clinicopathological Factors				
Histology				<0.0001
Ductal	12,814 (82%)	4,328 (34%)	8,486 (66%)	
Lobular	830 (5%)	114 (14%)	716 (86%)	
Mixed ductal and lobular	574 (4%)	103 (18%)	471 (82%)	
Other	1,335 (9%)	370 (28%)	965 (72%)	
Tumor Grade				<0.0001
1	864 (6%)	151 (17%)	713 (83%)	
2	5,172 (33%)	1,185 (23%)	3,987 (77%)	
3	9,517 (61%)	3,579 (38%)	5,938 (62%)	
Molecular Type				<0.0001
HR+/Her2-	6,675 (43%)	1,215 (18%)	5,460 (82%)	
HR+/Her2+	2,637 (17%)	1,033 (39%)	1,604 (61%)	
HR-/Her2+	1,746 (11%)	871 (50%)	875 (50%)	
HR-/Her2-	4,495 (29%)	1,796 (40%)	2,699 (60%)	
Clinical T Stage				<0.0001
T1	2,122 (14%)	865 (41%)	1,257 (59%)	
T2	7,103 (46%)	2,455 (35%)	4,648 (65%)	
T3	3,483 (22%)	869 (25%)	2,614 (75%)	
T4	2,845 (18%)	726 (26%)	2,119 (74%)	
Clinical N Stage				<0.0001
N0	5,654 (36%)	2,029 (36%)	3,625 (64%)	
N1	7,259 (47%)	2,174 (30%)	5,085 (70%)	
N2	1,630 (10%)	432 (27%)	1,198 (74%)	
N3	1,010 (6%)	280 (28%)	730 (72%)	

Table 2. Predictors of Pathological Complete Response-Multivariate Logistic Regression

	N	OR	95% CI
Age			
<30	208	2.11	1.30-3.44
30-39	1,605	2.05	1.37-3.07
40-49	3,447	1.92	1.30-2.86
50-59	3,798	1.69	1.14-2.50
60-69	2,469	1.43	0.96-2.13
70-79	760	1.32	0.86-2.01
80+	155	Reference	-
Race			
White	9,636	1.18	1.06-1.32
Black	2,134	Reference	-
Other	672	1.15	0.94-1.40
Histology			
Ductal	10,264	1.63	1.27-2.09
Lobular	641	Reference	-
Mixed ductal and lobular	460	1.24	0.88-1.74
Other	1,077	1.35	1.02-1.80
Tumor Grade			
1	699	Reference	-
2	4,134	1.05	0.85-1.30
3	7,609	1.78	1.44-2.21
Molecular Type			
HR+/Her2-	5,353	Reference	-
HR+/Her2+	2,120	2.60	2.32-2.92
HR-/Her2+	1,411	4.00	3.51-4.56
HR-/Her2-	3,558	2.37	2.13-2.63
Clinical T Stage			
T1	1,704	2.08	1.81-2.38
T2	5,667	1.47	1.32-1.63
T3	2,806	Reference	-
T4	2,265	1.05	0.85-1.30
Clinical N Stage			
N0	4,549	1.45	1.21-1.73
N1	5,780	1.18	0.99-1.40
N2	1,312	1.07	0.87-1.31
N3	801	Reference	-

Fig 1. Sample selection method.

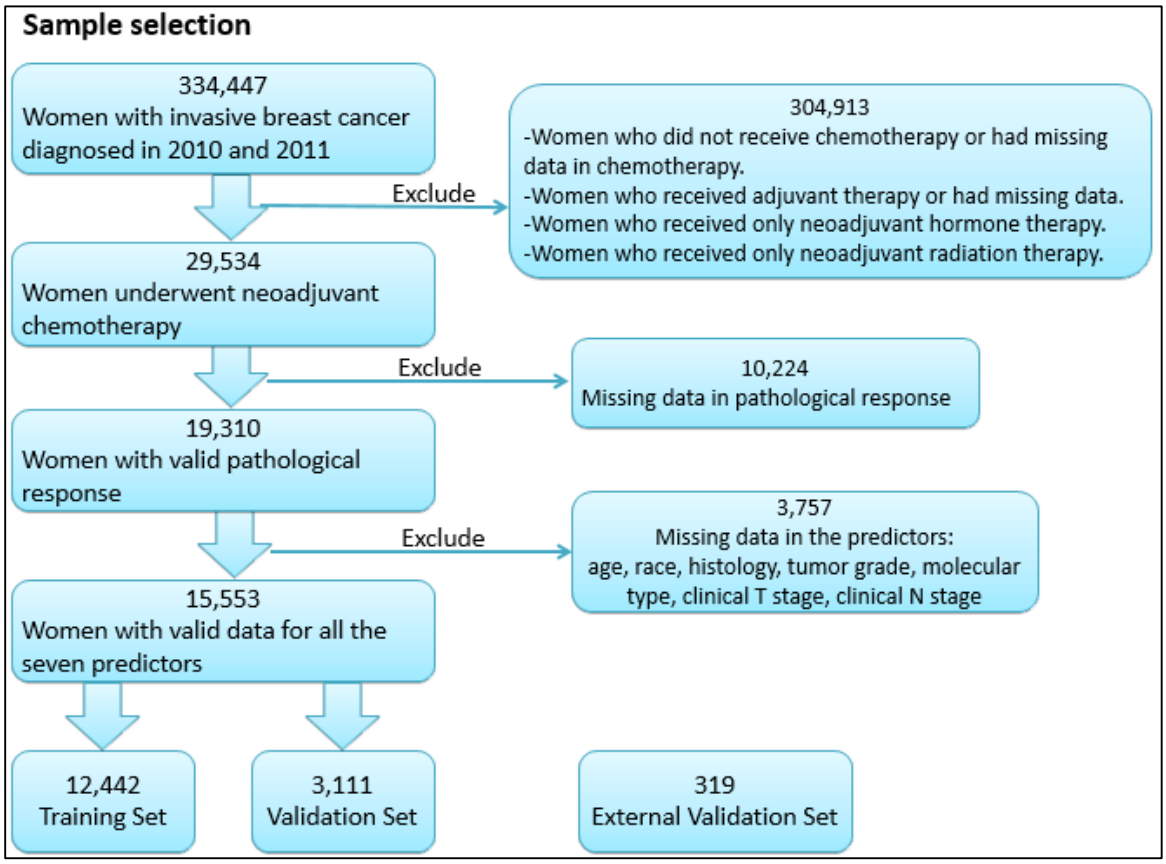


Fig 2. Nomogram to predict the probability of pathological complete response (pCR) to neoadjuvant chemotherapy.

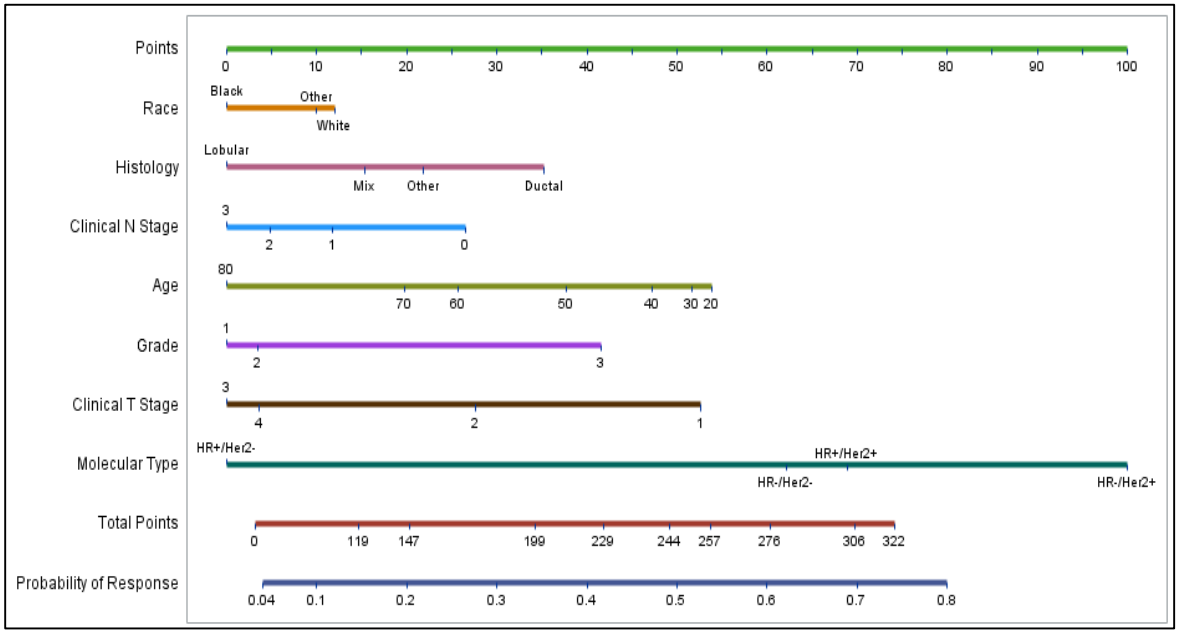
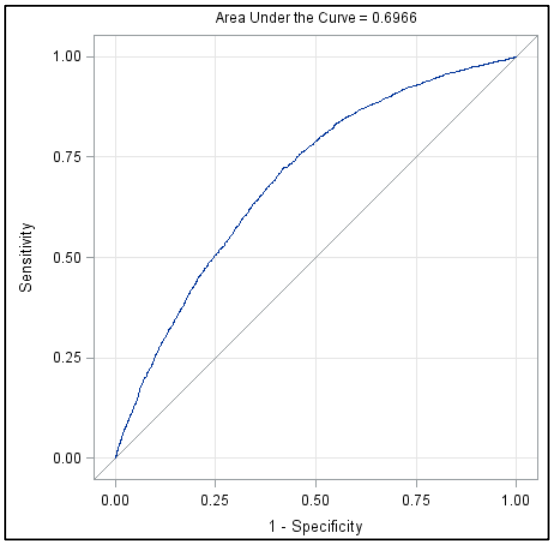
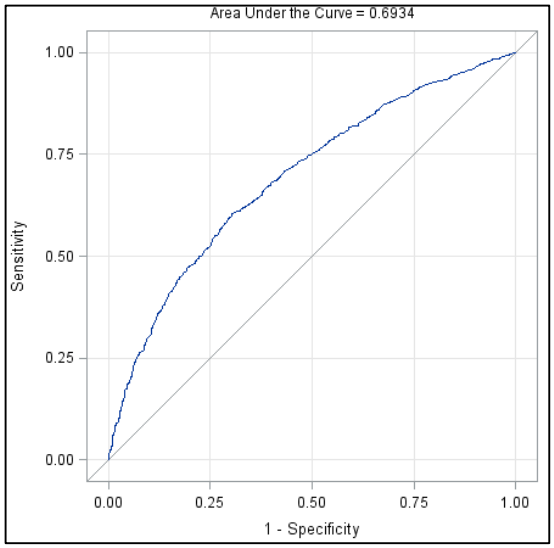


Fig 3. The ROC curves of prediction model in the A. training set, B. internal validation set, and C. external validation set.

A.



B.



C.

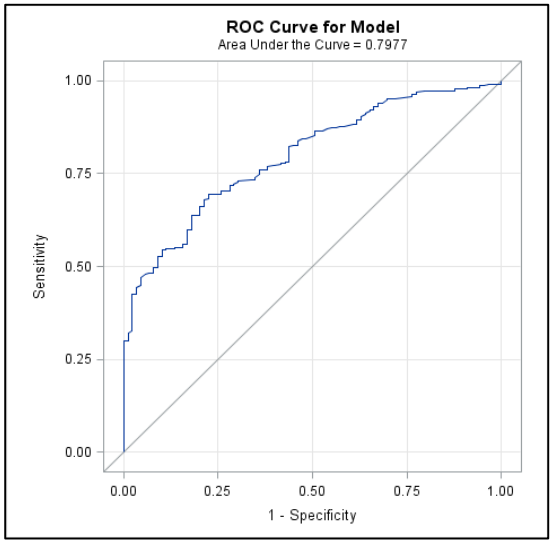
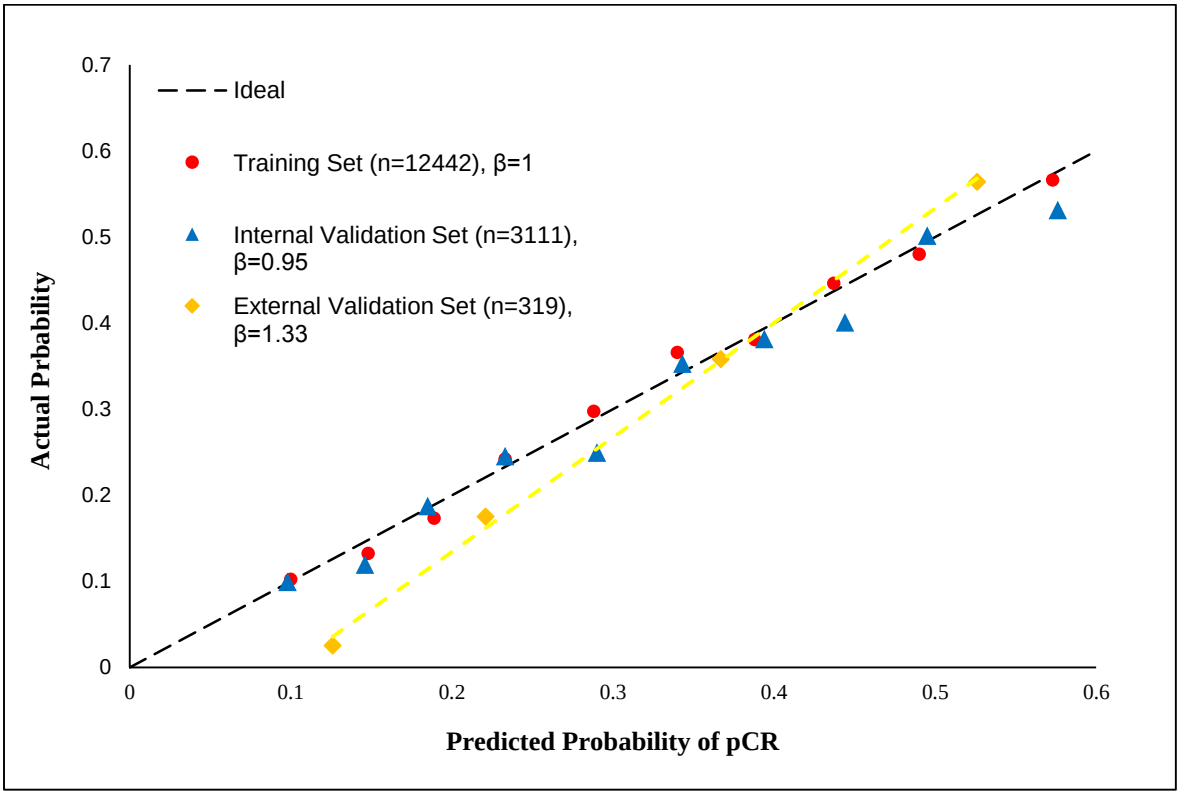


Fig 4. Calibration plot of the nomogram for probability pathological complete response (pCR) to neoadjuvant chemotherapy.



Reference

[1] DeSantis C, Siegel R, Bandi P, et al. Breast cancer statistics, 2011. *CA Cancer J Clin* 2011;61:409-18.

[2] Rouzier R, Pusztai L, Delaloge S, Gonzalez-Angulo AM, Andre F, Hess KR, Buzdar AU, Garbay JR, Spielmann M, Mathieu MC, Symmans WF, Wagner P, Atallah D, Valero V, Berry DA, Hortobagyi GN: Nomograms to predict pathologic complete response and metastasis-free survival after preoperative chemotherapy for breast cancer. *J Clin Oncol* 2005, 23:8331-8339.

[3] Mauri D, Pavlidis N, Ioannidis JP: Neoadjuvant versus adjuvant systemic treatment in breast cancer: A meta-analysis. J Natl Cancer Inst 97:188-194, 2005.

[4] Rouzier R, Mathieu MC, Sideris L, et al: Breast-conserving surgery after neoadjuvant anthracycline-based chemotherapy for large breast tumors. Cancer 101:918-925, 2004.

[5] Boughey JC1, McCall LM, Ballman KV, Mittendorf EA, Ahrendt GM, Wilke LG, Taback B, Leitch AM, Flippo-Morton T, Hunt KK. Tumor biology correlates with rates of breast-conserving surgery and pathologic complete response after neoadjuvant chemotherapy for breast cancer: findings from the ACOSOG Z1071 (Alliance) Prospective Multicenter Clinical Trial. Ann Surg. 2014, 260(4):608-14.

[6] Guidance for Industry Pathological Complete Response in Neoadjuvant Treatment of High-Risk Early-Stage Breast Cancer: Use as an Endpoint to Support Accelerated Approval. 2014, page 2-7.
http://www.fda.gov/downloads/drugs/guidancecomplianceregulatoryinformation/guidances/ucm305501.pdf.

[7] Rouzier R, Extra JM, Klijanienko J, et al: Incidence and prognostic significance of complete axillary downstaging after primary chemotherapy in breast cancer patients with T1 to T3 tumors and cytologically proven axillary metastatic lymph nodes. J Clin Oncol 20:1304-1310, 2002.

[8] Kuerer HM, Newman LA, Smith TL, et al: Clinical course of breast cancer patients with complete pathologic primary tumor and axillary lymph node response to doxorubicin-based neoadjuvant chemotherapy. J Clin Oncol 17:460- 469, 1999.

[9] Sparano JA. TAILORx: trial assigning individualized options for treatment (Rx). Clin Breast Cancer 7(4):347-50, 2006.

[10] F. Cardoso, M. Piccart-Gebhart, L Van't Veer, E. Rutgers. The MINDACT trial: the first prospective clinical validation of a genomic tool. Mol Oncol, 1(3): 246–252, 2007.

[11] Keam B, Im SA, Park S, Nam BH, Han SW, Oh DY, Kim JH, Lee SH, Han W, Kim DW, Kim TY, Park IA, Noh DY, Heo DS, Bang YJ: Nomogram predicting clinical outcomes in breast cancer patients treated with neoadjuvant chemotherapy. J Cancer Res Clin Oncol 2011, 137:1301-1308.

[12] Colleoni M, Bagnardi V, Rotmensz N, Dellapasqua S, Viale G, Pruneri G, Veronesi P, Torrisi R, Luini A, Intra M, Galimberti V, Montagna E, Goldhirsch A. A risk score to predict disease-free survival in patients not achieving a pathological complete remission after preoperative chemotherapy for breast cancer. Ann Oncol 20:1178–1184, 2009.

[13] Commission on Cancer. Data acquisition manual. Chicago, IL: American College of Surgeons, 1990.

[14] BeahrsO, HensonDE, HutterRVP, KennedyBJ, editors. Manual for staging of cancer. 4th ed. Philadelphia: J.B. Lippincott Co., 1992.

[15] ICD-O. International classification of disease for oncology. 2nd edition. Geneva: World Health Organization, 1990.

[16] Howe, J. R., Karnell, L. H., Menck, H. R. and Scott-Conner, C. Adenocarcinoma of the small bowel. Cancer, 86: 2693–2706, 1999.

[17] 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(1–2):4–13.

[18] Perou CM, Sørlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA, Fluge O, Pergamenschikov A, Williams C, Zhu SX, Lønning PE, Børresen-Dale AL, Brown PO, Botstein D. Molecular portraits of human breast tumours. Nature. 2000 Aug 17; 406(6797):747-52

[19] Sotiriou C, Neo SY, McShane LM, Korn EL, Long PM, Jazaeri A, Martiat P, Fox SB, Harris AL, Liu ET. Breast cancer classification and prognosis based on gene expression profiles from a population-based study. Proc Natl Acad Sci U S A. 2003 Sep 2; 100(18):10393-8.

[20] Sorlie T, Tibshirani R, Parker J, et al. Repeated observation of breast tumor subtypes in independent gene expression data sets. Proc Natl Acad Sci U S A 2003;100:8418-23.

[21] Pickard JM. Using receiver operating characteristic (ROC) curves to evaluate digital mammography. University of Texas Health Science Center.

[22] Hanley, JA, McNeil, BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 1982, 143, 29-36.

[23] Cortazar P. et al. Pathological complete response and long-term clinical benefit in breast cancer: the CTNeoBC pooled analysis. Lancet. 384: 164–172, 2014.

[24] Fisher B, Bryant J, Wolmark N, et al: Effect of preoperative chemotherapy on outcome of women with operable breast cancer. J Clin Oncol 16:2672-2685, 1998.

[25] Rouzier R, Pusztai L, Garbay JR, Delaloge S, Hunt KK, Hortobagyi GN, Berry D, Kuerer HM (2006) Development and validation of nomograms for predicting residual tumor size and the probability of successful conservative surgery with neoadjuvant chemotherapy for breast cancer. Cancer 107:1459–1466.

[26] Criscitiello C, Azim HA, Agbor-Tarh D, et al: The discrepancy between high pathological complete response rate and low breast conserving surgery following neoadjuvant therapy: Analysis from the NeoALLTO trial (BIG 1-06). 2012 ESMO Congress. Abstract 2470. Presented September 30, 2012.

[27] Baselga J, Bradbury I, Eidtmann H, et al: Lapatinib with trastuzumab for HER2-positive early breast cancer (NeoALTTO): A randomised, open-label, multicentre, phase 3 trial. Lancet 379:633-640, 2012.

[28] Yip CH (2009) Breast cancer in Asia. Methods Mol Biol 471:51–64.

[29] Yoo KY, Kang D, Park SK, Kim SU, Shin A, Yoon H, Ahn SH, Noh DY, Choe KJ. Epidemiology of breast cancer in Korea: occurrence, high-risk groups, and prevention. J Korean Med Sci 17:1–6, 2002.

[30] Guarneri V, Broglio K, Kau S-W, Cristofanilli M, Buzdar AU, Valero V, Buchholz T, Meric F, Middleton L, Hortobagyi GN, Gonzalez-Angulo AM. Prognostic value of pathologic complete response after primary chemotherapy in relation to hormone receptor status and other factors. J Clin Oncol 2006, 24(7):1037-1044.

[31] Killelea BK, Yang VQ, Mougalian S, Horowitz NR, Pusztai L, Chagpar AB, Lannin DR. Neoadjuvant Chemotherapy for Breast Cancer Increases the Rate of Breast Conservation: Results from the National Cancer Database. J Am Coll Surg. 2015 Feb 26. pii: S1072-7515(15)00145-3. doi: 10.1016/j.jamcollsurg.2015.02.011. [Epub ahead of print].