The Breast 46 (2019) 116–125

Contents lists available at ScienceDirect

The Breast

journal homepage: www.elsevier.com/brst

Original article

Nomogram update based on TAILORx clinical trial results - Oncotype DX breast cancer recurrence score can be predicted using clinicopathologic data☆,☆☆

Amila Orucevic [a, *], John L. Bell [b], Megan King [c], Alison P. McNabb [d], Robert E. Heidel [b]

[a] Department of Pathology, The University of Tennessee Medical Center, 1924 Alcoa Hwy, Knoxville, TN, 37920, USA
[b] Department of Surgery, The University of Tennessee Medical Center, 1924 Alcoa Hwy, Knoxville, TN, 37920, USA
[c] College of Medicine, The University of Tennessee Health Science Center, 910 Madison Avenue, Ste 1031, Memphis, TN, 38163, USA
[d] Graduate School of Medicine, The University of Tennessee Medical Center, 1924 Alcoa Hwy, Knoxville, TN, 37920, USA

ARTICLE INFO

Article history:
Received 16 December 2018
Received in revised form
30 April 2019
Accepted 2 May 2019
Available online 10 May 2019

Keywords:
Invasive breast cancer
Oncotype DX
Recurrence score prediction
Clinicopathologic variables
Breast cancer nomograms
On-line nomogram
Calculator for prediction of Oncotype DX recurrence score

ABSTRACT

Objectives: Oncotype DX (ODX), 21-gene breast cancer (BC) assay, predicts risk of recurrence and benefits of addition of chemotherapy to hormonal therapy for early-stage BC. We previously published a nomogram/calculator that could predict ODX results without performing the test by using clinicopathologic characteristics of BC available from pathology reports. Patients with intermediate-risk (11–25) ODXRS (RS) were excluded from that nomogram. This update tests the predictive value of clinicopathologic variables for forecasting the ODXRS while including intermediate-risk-ODXRS patients and stratifying ODXRS based on recently published TAILORx clinical trial results (0–25 = low-risk, 26–100 = high-risk-ODXRS; intermediate-risk-ODXRS belongs to the low-risk category).

Material and methods: The nomogram was built on 65,754 ODX-tested ER+/HER2-/lymph-node-negative patients with 6–50 mm tumor, captured by the National Cancer Data Base (NCDB) from 2010 to 2014. Five clinicopathologic variables (age, tumor size, grade, progesterone-receptor status (PR) and BC-histologic type) were assessed with logistic regression to predict for a low-risk (0–25) or a high-risk (26–100) ODXRS. Results were validated on a separate 18,585 ODX-tested cohort from 2015.

Results: Grade and PR were the highest significant predictors of both low-risk and high-risk-ODXRS, followed by histologic type, tumor size and age. The Receiver Operator Characteristic (ROC) curve showed strong statistical model for both low-risk and high-risk-ODXRS prediction outcomes (c-index = 0.81).

Conclusions: An updated nomogram is now developed/validated on the entire population of ODX-tested patients (84,339) captured by the NCDB. The nomogram/calculator, available on-line at the UTMCK/Shiny website (https://utgsm.shinyapps.io/OncotypeDXCalculator/), will continue serving as a surrogate for BC patients for which ODX testing is not affordable, available or necessary.

Abbreviations: ODX, Oncotype DX; BC, breast cancer; RS, recurrence score; ODXRS, Oncotype DX recurrence score; NCDB, National Cancer Data Base; ER, estrogen receptor; HER2, human epidermal growth factor receptor 2; PR, progesterone receptor; ROC curve, Receiver Operator Characteristics curve; SD, standard deviation; IDC, invasive ductal carcinoma; ILC, invasive lobular carcinoma; IDC + ILC, invasive ductal and lobular carcinoma; IDC + others, invasive ductal carcinoma mixed with other types; AOR, adjusted odds ratio; 95% CI, 95% confidence interval; β, beta coefficient; PPV, positive predictive value; NPV, negative predictive value.

☆ **Note:** Part of this study was presented at the San Antonio Breast Cancer Symposium in the poster session (P2-08-09) on December 6, 2018.

☆☆ **Disclaimers:** The National Cancer Data Base (NCDB) is a joint project of the Commission on Cancer (CoC) of the American College of Surgeons and the American Cancer Society. The CoC's NCDB and the hospitals participating in the CoC NCDB are the source of the de-identified data used herein; they have not verified and are not responsible for the statistical validity of the data analysis or the conclusions derived by the authors.

* Corresponding author. The University of Tennessee, Medical Center at Knoxville, Pathology, 1924 Alcoa Hwy, Knoxville, TN, 37920, United States.
E-mail address: aorucevic@utmck.edu (A. Orucevic).

https://doi.org/10.1016/j.breast.2019.05.006

Introduction

Oncotype DX (ODX), a commercially available 21-gene recurrence score breast cancer assay (Genomic Health, Redwood City, CA), provides prognostic and predictive breast cancer recurrence information for hormone-positive, HER2-negative, lymph-node negative patients [1,2]. It is endorsed by the National Comprehensive Cancer Network (NCCN) [3] and the American Society of Clinical Oncology (ASCO) [4]. It is used for staging purposes by the revised 8th edition (2018) of American Joint Commission on Cancer staging manual [5].

Initial results from the TAILORx clinical trial published in 2015 [6] prospectively validated a 21-gene expression assay in breast cancer. The study revealed that patients with low Oncotype DX recurrence score (ODXRS) from 0 to 10 can safely forego adjuvant chemotherapy, while adjuvant chemotherapy is recommended for high ODXRS (≥26). No adjuvant chemotherapy recommendations were previously available for intermediate (11–25) ODXRS patients until 6/3/2018, when the TAILORx clinical trial results on intermediate ODXRS were published [7]. According to the new data, patients with ODXRS 11–25 can now safely forego adjuvant chemotherapy, although some benefit of adjuvant chemotherapy was noted in patients younger than 50 years of age with ODXRS of 16–25. These new data allow categorization of ODXRS as a binary variable (0–25 as a low-risk ODXRS and 26–100 as a high-risk ODXRS), in which intermediate-risk score patients (11–25) are now included in the low-risk category.

According to Orucevic's et al. data, ODX is performed in approximately 1/3 of eligible early-stage breast cancer patients in the United States [8]. It is performed for less than 20% of patients in European countries [9]. ODX use in the United States is greatly influenced by race and socioeconomic status. Patients of Caucasian race and high socioeconomic status are tested significantly more often than patients of other races and low socioeconomic status [8,10–13].

We previously developed and published [14] a user-friendly nomogram based on clinicopathologic characteristics of ODX tested patients captured by the National Cancer Data Base (NCDB) as a surrogate prediction model for the ODX assay. Intermediate-risk ODXRS patients (11–25) were excluded from that nomogram. The objective of this update is to test the predictive value of clinicopathologic variables for forecasting the new TAILORx binary ODXRS stratification, in which intermediate-risk ODXRS patients are included in the low-risk category (0–25 = low-risk). Five clinicopathologic variables readily available from pathology reports and established in clinical practice as prognostic and/or predictive were chosen for the development of the nomogram: age, tumor size, tumor grade, progesterone receptor status (PR) and the histologic type of breast cancer. Lymphovascular invasion was excluded in the current nomogram since lymphovascular invasion is now deemed as an optional data reporting element for invasive breast cancer by the College of American Pathologists' guidelines [15].

Materials and methods

Patients and pathology variables selection

Nomograms that predict for a low-risk or a high-risk ODXRS (outcomes of interest) were constructed based on the methods published by Iasonos et al. [16]. Low-risk and a high-risk ODXRS cut-off values were based on the recently published TAILORx clinical trial results [7] (0–25 = low-risk ODXRS and 26–100 = high-risk ODXRS).

Some benefit of adjuvant chemotherapy was identified in the TAILORx trial [7] in patients younger than 50 years of age with ODXRS of 16–25. Therefore, we also evaluated our prediction model for ODXRS cut-off values 0–15 for a low-risk and 16–100 for a high-risk ODXRS.

ODX-tested breast cancer patients captured by NCDB from 2010 to 2014, with results recorded as a numerical value (0–100), served as the training study cohort. A validation study cohort was chosen from a separate sample of ODX-tested breast cancer patients collected by the NCDB in 2015.

The NCDB is a clinical oncology database which gathers data from cancer registries from more than 1500 Commission on Cancer-accredited facilities. NCDB captures more than 70% of all newly diagnosed malignancies in the United States [17]. Prior to the release of the data files, the NCDB de-identifies data that include the names of patients and institutions; therefore, criteria of 45 CFR 46.102 d research are met, and Institutional Review Board approval is not required.

Inclusion criteria for creation of the current nomogram were similar to the previously published nomogram [14]: 1) female, 2) the four most frequent histologic types of invasive breast carcinoma (invasive ductal, invasive lobular, invasive ductal and lobular, and invasive ductal carcinoma mixed with other types), 3) ER positive (≥1%), 4) HER2 negative, 5) no regional lymph node metastasis, 6) tumor size between 6 mm and 50 mm, 7) ODX test results recorded as a numerical value (0–100).

Statistical methods and nomogram development

The methods published by Iasonos et al. [16] and used for construction of our previous nomograms [14] were employed in creation of the updated nomograms. Recommendations on transparent reporting of a multivariable prediction model for individual prognosis or diagnosis from the TRIPOD Statement [18] were followed.

Development of the training model

The logistic regression model was created with tumor size, tumor grade, PR status, histologic type of breast cancer and age as predictor variables. A low-risk or a high-risk ODXRS were the outcome variables. Multicollinearity was assessed using the same methods as previously described [14,16] (detailed in the summary of statistical analysis, supplemental data). Discrimination of the training model, i.e the ability to differentiate between those who did or did not have the observed outcome was assessed using a receiver operator characteristic (ROC) curve. Discrimination was estimated by the concordance index (c-index) with 95% confidence interval (95% CI) Calibration, the agreement between outcome predictions and the observed outcomes was also assessed.

Internal validation of the training model

Internal validation of the training model was performed using the bootstrap validation approach on 1000 cases (detailed in Supplementary data, Summary of statistical analysis)

External validation of the training model and model performance (detailed in the Supplemental data)

The predictive performance of the training model was assessed in the external validation cohort in regards to both discrimination and calibration. Logistic regression formula obtained from the training cohort was applied to the external validation cohort. The ROC curve and c-index were then compared with the training cohort.

Additional predictive performance measures of the model such

Table 1
Clinicopathologic characteristics of a training and a validation cohort patients used for building of nomograms to predict for a low-risk (0–25) and a high-risk (26–100) Oncotype DX recurrence score.

Clinico-pathologic characteristic	Training cohort NCDB 2010–2014			Validation cohort NCDB 2015		
	Total No (% total)	ODX Low-risk (0–25) No (% total)	ODX High-risk (26–100) No (% total)	Total No (% total)	ODX Low-risk (0–25) No (% total)	ODX High-risk (26–100) No (% total)
Sex - Female	65,754 (100)	55,651 (84.6)	10,103 (15.4)	18,585 (100)	15,839 (85.2)	2746 (14.8)
Age (19–90) (mean ±SD)	65,754 58 ± 10.3 (100)	55,651 (84.6)	10,103 (15.4)	18,585 58.7 ± 10.2 (100)	15,839 (85.2)	2746 (14.8)
≤50	16,844 (25.6)	14,433 (21.9)	2411 (3.7)	4331 (23.3)	3743 (20.1)	588 (3.2)
>50	48,910 (74.4)	41,218 (62.7)	7692 (11.7)	14,254 (76.7)	12,096 (65.1)	2158 (11.6)
Tumor size (6–50 mm)	65,754 (100)	55,651 (84.6)	10,103 (15.4)	18,585 (100)	15,839 (85.2)	2746 (14.8)
Mean ± SD	16.1 ± 7.5			16.1 ± 7.7		
T1b	15,436 (23.5)	13,661 (20.8)	1775 (2.7)	4423 (23.8)	3920 (21.1)	503 (2.7)
T1c	35,956 (54.7)	30,691 (46.7)	5265 (8.0)	9966 (53.6)	8592 (46.2)	1374 (7.4)
T2	14,362 (21.8)	11,299 (17.2)	3063 (4.7)	4196 (22.6)	3327 (17.9)	869 (4.7)
Histologic type	65,754 (100)	55,651 (84.6)	10,103 (15.4)	18,585 (100)	15,839 (85.2)	2746 (14.8)
IDC	51,931 (79)	43,013 (65.4)	8918 (13.6)	14,401 (77.5)	12,011 (64.6)	2390 (12.9)
ILC	7051 (10.7)	6558 (10.0)	493 (0.7)	2234 (12)	2068 (11.1)	166 (0.9)
IDC + ILC	4480 (6.8)	4.057 (6.2)	423 (0.6)	1355 (7.3)	1241 (6.7)	114 (0.6)
IDC + others	2292 (3.5)	2023 (3.1)	269 (0.4)	595 (3.2)	519 (2.8)	76 (0.4)
Histologic grade	65,754 (100)	55,651 (84.6)	10,103 (15.4)	18,585 (100)	15,839 (85.3)	2746 (14.8)
Grade 1	18,243 (27.7)	17,498 (26.6)	745 (1.1)	5478 (29.5)	5246 (28.2)	232 (1.2)
Grade 2	36,297 (55.2)	32,191 (49)	4106 (6.2)	10,137 (54.5)	9026 (48.6)	1111 (6.0)
Grade 3	11,214 (17.1)	5962 (9.1)	5252 (8.0)	2970 (16)	1567 (8.4)	1403 (7.5)
Progesterone receptor	65,754 (100)	55,651 (84.6)	10,103 (15.4)	18,585 (100)	15,839 (85.2)	2746 (14.8)
Positive	59,578 (90.6)	52,503 (79.8)	7075 (10.8)	16,873 (90.8)	14,970 (80.5)	1903 (10.2)
Negative	6176 (9.4)	3148 (4.8)	3028 (4.6)	1712 (9.2)	869 (4.7)	843 (4.5)

ODX – Oncotype DX; SD – standard deviation; IDC – invasive ductal carcinoma; ILC – invasive lobular carcinoma; IDC + ILC – invasive ductal and lobular carcinoma; IDC + others – invasive ductal carcinoma mixed with other types.

as Nagelkerke R^2, sensitivity, specificity, positive predictive value, negative predictive value were also reported.

Creation of the nomogram

The nomograms were created on the training model cohort (65,754 patients); methodology is detailed in Supplementary data.

All statistical analyses were conducted using SPSS Version 25 (Armonk, NY: IBM Corp.). The analyses were then confirmed in R program (Vienna, Austria) [19–21]. Details of analysis are presented in the Supplementary data. An online nomogram/calculator was developed using Shiny application, an open source R package. The calculator is available at the following website: https://utgsm.shinyapps.io/OncotypeDXCalculator/

Results

Clinicopathologic characteristics of patients in training and validation cohorts used for building of nomograms/calculator

The training cohort consisted of 65,754 ODX tested ER+/HER2-/lymph-node-negative patients with 6–50 mm tumor size, captured by the NCDB from 2010 to 2014. An external validation cohort consisted of 18,585 ODX tested patients captured by NCDB in 2015. Descriptive clinicopathologic characteristics of these patients used for building of nomograms were presented in Table 1.

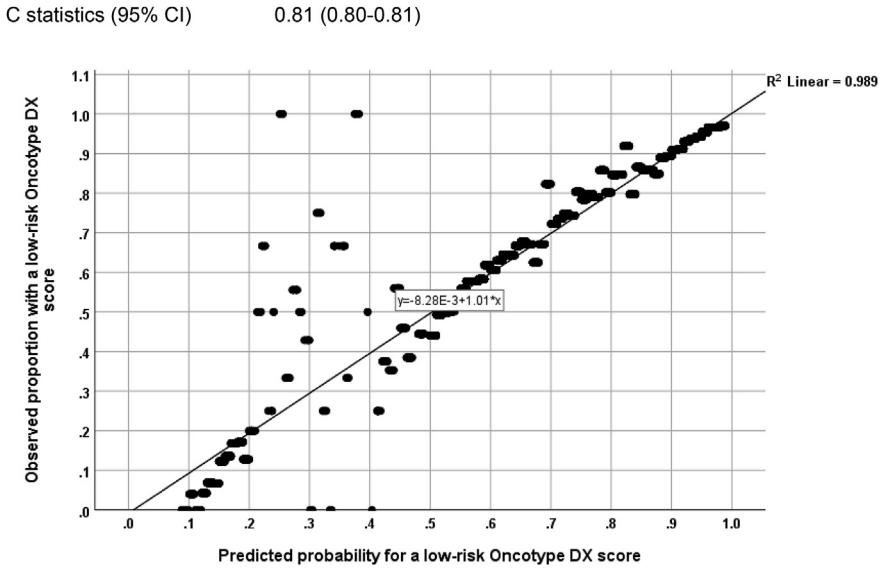

Fig. 1. Calibration plot: Model predicted probability vs. observed proportion of patients with a low-risk ODX score (training cohort).

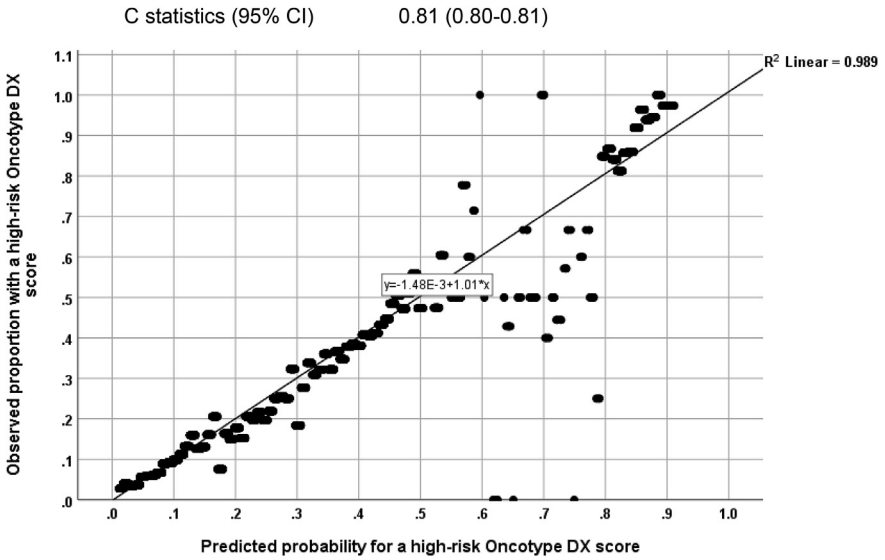

Fig. 2. Calibration plot: Model predicted probability vs observed proportion of patients with a high-risk ODX score (training cohort).

Nomogram/calculator model development

In multivariate logistic regression analysis, all five variables, tumor size, tumor grade, PR status, histologic tumor type, and age, were significantly associated with a low-risk or a high-risk ODXRS test result. Results of multivariate logistic regression analysis for a low-risk and a high-risk ODXRS prediction including β values (regression coefficients) for five variables and logistic regression constant (the model intercept) were presented in Supplementary Tables 1 and 2, respectively. Tumor grade and PR status had the highest significant impact on predicting a high-risk or a low-risk ODXRS, followed by the histologic tumor type, tumor size, and age. For example, grade 3 tumor was 18.36 times more likely to be associated with a high-risk ODXRS than a grade 1 tumor (95% CI 16.84–20.01, $p < .001$); negative PR status was 7.62 times more likely to be associated with a high-risk ODXRS than a positive PR status (95% CI 7.14–8.14, $p < .001$; Supplementary Table 2).

Performance of the model: Discrimination, accuracy, sensitivity, specificity, calibration

No evidence of multicollinearity was found. The ROC curve or C statistics was used to measure the discriminatory capacity of our model to predict for a low-risk (0–25) and a high-risk (26–100) ODXRS test result, respectively. The c-index for both was 0.81 (95% CI 0.80–0.81), indicating a strong prediction model [22,23] (Supplementary Fig. 1).

The overall accuracy was 86.8% (model correctly predicted observed ODX risk score (low or high) in 86.8% of cases). When predicting for a low-risk ODX score, sensitivity was 99.2%, specificity 18.6%, PPV 87% and NPV 80.2%. When predicting for a high-risk ODX score, sensitivity was 18.6%, specificity 99.2%, PPV 80.2% and NPV 87% (Supplementary Tables 3 and 4).

Model calibration (accuracy of prediction) was assessed with calibration plots (Figs. 1 and 2).

Fig. 1 showed an excellent calibration when predicting for a low-

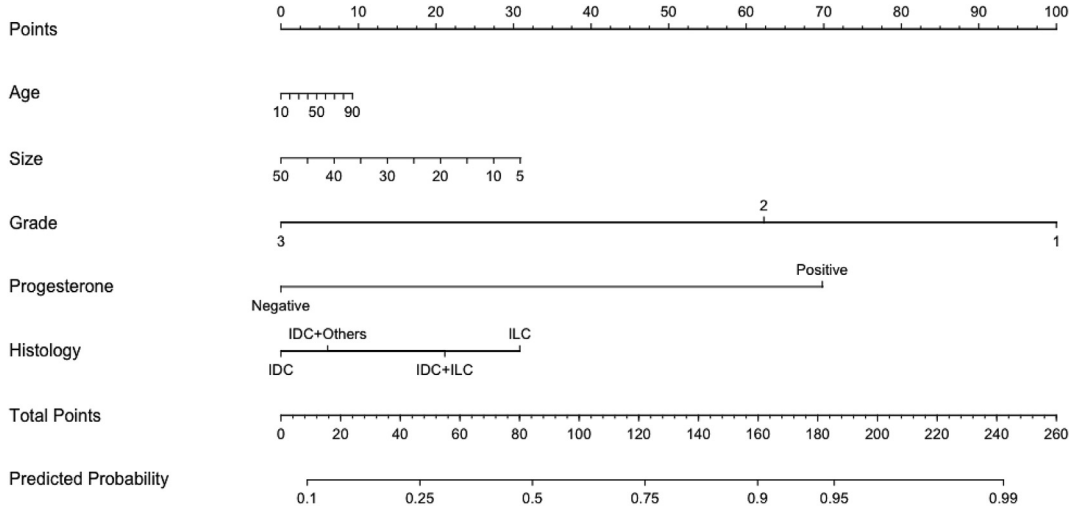

Fig. 3. Nomogram to predict for a **low-risk** Oncotype DX (ODX) score (TAILORx clinical trial cut-off values 0–25). Size – tumor size in mm; Progesterone – progesterone receptor status; Histology – histologic type of breast cancer; IDC – invasive ductal carcinoma; ILC – invasive lobular carcinoma; IDC + ILC – invasive ductal and lobular carcinoma; IDC + others – invasive ductal carcinoma mixed with other types; Predicted probability – predicted probability for a low-risk (0–25) Oncotype DX recurrence score.

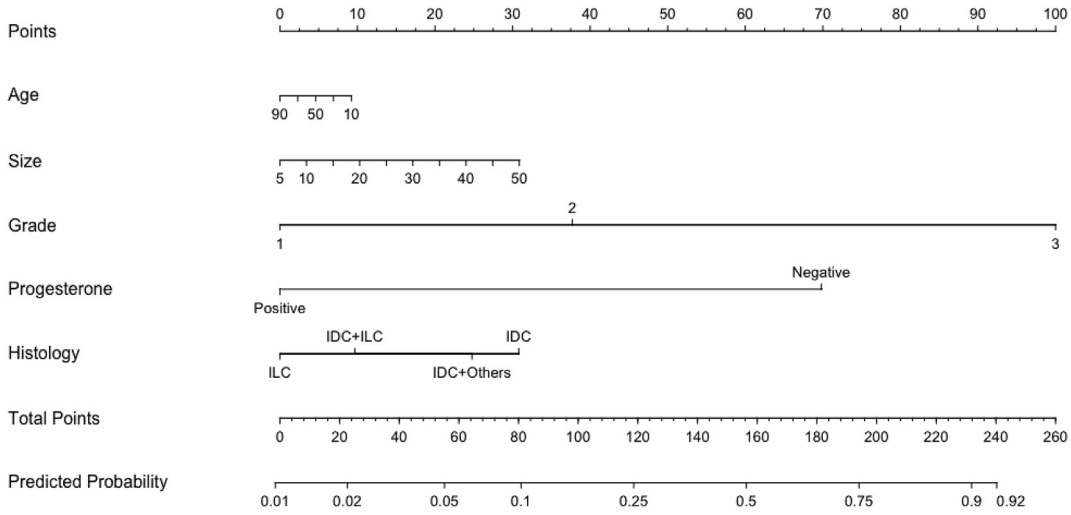

Fig. 4. Nomogram to predict for a **high-risk** Oncotype DX (ODX) score (TAILORx clinical trial cut-off values 26–100). Size – tumor size in mm; Progesterone – progesterone receptor status; Histology – histologic type of breast cancer; IDC – invasive ductal carcinoma; ILC – invasive lobular carcinoma; IDC + ILC – invasive ductal and lobular carcinoma; IDC + others – invasive ductal carcinoma mixed with other types; Predicted probability – predicted probability for a high-risk (26–100) Oncotype DX recurrence score.

risk ODX score with an R^2 (linear regression) of 0.989. Predictions and the observed outcomes were almost identical from 50% to 100% as represented by the proximity of the dots to the regression line. Our regression line y = 1.01x - 0.008 was very close to the ideal prediction line y = x.

Fig. 2 is a mirror image of Fig. 1.

Some benefit of adjuvant chemotherapy was identified in the TailorX clinical trial [7] in patients younger than 50 years of age with ODXRS of 16–25. Therefore, we also evaluated our prediction model for ODXRS cut-off values 0–15 for a low-risk and 16–100 for a high-risk ODXRS. A c-index of 0.676 (96%CI 0.672–0.68) was calculated for these cut-off values. As this c-index was regarded as an indicator of a poor discrimination model [22,23], these cut-off values were not pursued for building of the nomograms.

Table 2
Comparison of nomogram/calculator calculated probability to observed proportion of patients with a low-risk ODX score at probability range from 85% to 100%, potentially relevant for clinical use. Total number of patients included in the table: 50,461 (training cohort).

Nomogram calculated probability for a low-risk ODX (%)	$\frac{\text{Observed \# of patients with low-risk ODX}}{\text{total \# of patients for a given probability}}$ (%)
85%	$\frac{145}{168}$ (86.3%)
86%	$\frac{243}{277}$ (87.7%)
87%	$\frac{481}{569}$ (84.5%)
88%	$\frac{1213}{1385}$ (87.6%)
89%	$\frac{2632}{3188}$ (88.8%)
90%	$\frac{6326}{7026}$ (90.0%)
91%	$\frac{9196}{10082}$ (91.2%)
92%	$\frac{3463}{3776}$ (91.7%)
93%	$\frac{568}{596}$ (95.3%)
94%	$\frac{1176}{1256}$ (93.6%)
95%	$\frac{2599}{2753}$ (94.4%)
96%	$\frac{5610}{5825}$ (96.3%)
97%	$\frac{10652}{11025}$ (96.6%)
98%	$\frac{1508}{1552}$ (97.2%)
99%	$\frac{952}{983}$ (96.8%)
100%	No observed cases

Results from the external validation: Discrimination, accuracy, sensitivity, specificity

For both low-risk and high-risk ODX score, the predictive quality of our model remained the same: the calibration plot showed an excellent calibration with R^2 of 0.959 (Supplemental Fig. 2 and 3), c-index remained the same at 0.812 [95% CI 0.803–0.822], as well as calculated sensitivity, specificity, PPV and NPV (Supplemental Fig 1-3; Supplemental Tables 3 and 4). These results validated our training model as a strong model for predicting for both low- and high-risk ODX test results.

Table 3
Comparison of nomogram calculated probability to observed proportion of patients with a high-risk ODX score at probability range from 85% to 100%, potentially relevant for clinical use. Total number of patients included in the table: 644 (training cohort).

Nomogram calculated probability for a high-risk ODX (%)	$\frac{\text{Observed \# of patients with high-risk ODX}}{\text{total \# of patients for a given probability}}$ (%)
85%	$\frac{237}{260}$ (91.2%)
86%	$\frac{148}{154}$ (96.1%)
87%	$\frac{109}{115}$ (94.8%)
88%	$\frac{62}{66}$ (93.9%)
89%	$\frac{23}{23}$ (100%)
90%	$\frac{15}{16}$ (93.8%)
91%	$\frac{10}{10}$ (100%)
>=92%	No observed cases

Using the nomogram graphs and on-line calculator

Nomograms were developed based on the analyses in the training cohort group (n = 65,754). Points assigned for each clinicopathologic variable in nomograms are shown in Supplemental Table 5.

There are eight rows with numerical lines in each nomogram: a numerical line for each of the 5 clinicopathologic variables, a 0–100 line (topmost line), total points line and a low-risk or a high-risk ODX score predicted probability line (Figs. 3 and 4).

Based on an individual patient's clinicopathologic tumor characteristics obtained from the pathology report, each of the five predicting clinicopathologic variables is assigned a numerical value as follows: a vertical line is drawn between the variable's numerical line and the topmost "points" line in order to get number of points for each clinicopathologic variable (see also Supplemental Table 5 for assigned points). All assigned points for five variables are added and the total number of points is then found in the "total points" row numerical line. A vertical line is subsequently drawn between the final "total points" line and the "predicted probability" numerical line, allotting the final predicted probability for a low-risk or a high-risk ODXRS for an individual patient.

Nomograms from the training cohort were used for development of our updated online nomogram/calculator. Online calculator is available on the following website: https://utgsm.shinyapps.io/OncotypeDXCalculator/. Source code for R used in programming of the calculator is presented in the Supplementary data.

Online nomogram/calculator calculates probability for a low-risk or a high-risk ODXRS with 95% confidence intervals. Prediction intervals based on 95% confidence bounds, assuming a gaussian distribution, are calculated with standard error from 65,754 ODX tested patients in the training cohort. Calculated probability for a low-risk or a high-risk ODXRS ranges from 0 to 100%.

Nomogram/Calculator probability ranges for a prediction of ODX test results potentially relevant for clinical practice

Prediction for a low-risk ODX test results: As shown on the calibration plot in Fig. 1, our model has the highest accuracy for the probabilities from 50% to 100%. Even though the probabilities within the whole 50–100% range are informative for clinicians, those in the upper part of the range might be relevant for clinical use instead of the ODX test itself. The Table 2 shows calculated probability vs observed outcome for 85%–100% range.

Prediction for a high-risk ODX test results: As for the low-risk ODX, 85%–100% predicted probability range is potentially the most relevant for clinicians (see Table 3).

Predictive ability of the training model for correct categorization of a low-risk or a high risk ODX test results are compared to the actual/observed low-risk or a high-risk ODX test results in the training cohort for probability range from 85% to 100%. Results are presented in Table 4.

Discussion

ODX continues to be the most commonly utilized breast cancer genomic assay in the United States based on the data gathered by the NCDB in the time period from 2010 to 2015. From 434,592 ER+/HER2-negative/lymph node-negative female breast cancer patients, 167,589 (38.5%) had genomic tests performed, with ODX being the most frequently utilized test (95.8%). While performance of breast cancer genomic assays gives personalized prognostic and predictive information for individual breast cancer patients, these tests are expensive with ODX current estimated cost of ~$4000 [24]. Cost is likely one of the major factors contributing to the rate of utilization of the test, since even in the most developed countries such as the United States, the test was performed in slightly over one-third of eligible patients and in less than 20% of patients in European countries [9]. We [8] and others [10–13] have shown that utilization of the ODX test in the United States was influenced by socioeconomic status and race.

For the past few years, several institution-based studies, restricted to limited number of patients from respective institutions, were published utilizing some of the pathologic variables available from pathology reports to predict ODX score [17,25–32] (summary of select studies presented in Table 5). Recently published studies implied that ordering of ODX test for every eligible ER+/HER2-negative/lymph node-negative female

Table 4
Predictive ability of the training model for correct categorization of a low-risk or a high-risk ODX test results compared to the actual/observed low-risk or a high-risk ODX test results in the training cohort for probability range from 85% to 100%. Accuracy, sensitivity, specificity, positive predictive value and negative predictive value are also presented.

TRAINING COHORT (Probability range 85%–100%) n = 51,105	*Predicted* ODX score ≤25	*Observed* Low-risk ODX score ≤25	
		Low	High
	Low	46,764	3,697
	High	40	604
	Overall accuracy	92.7%	
	Sensitivity	99.9%	
	Specificity	14%	
	Positive predictive value	92.6%	
	Negative predictive value	94.3%	

TRAINING COHORT (Probability range 85%-100%) n = 51,105	*Predicted* ODX score >25	*Observed* High-risk ODX score >25	
		High	Low
	High	604	40
	Low	3,697	46,764
	Overall accuracy	92.7%	
	Sensitivity	14%	
	Specificity	99.9%	
	Positive predictive value	94.3%	
	Negative predictive value	92.6%	

Table 5
Summary of selected published studies that used sets of different clinicopathologic variables to predict ODX test results and comparison to our current study.

		Klein et al., 2013 UPMC [30]	Gage et al., 2015 AAMC [26]	Kim et al., 2016 JH [29]	Eaton et al., 2017 MSKCC [38]	Orucevic et al., 2017 UTMC [14]	Orucevic et al. Current study
Patients	Training	817	221	1113	766	27,719	65,754
	Validation	255	319 + 108	472	299	12,763	18,585
Age	Range, Mean ± SD	Not reported	Not reported	3-tier <50 50–69 >70	24-90 58 (median)	20-90 58.4 ± 10.3	19-90, 58 ± 10.3
Clinico-pathologic variables used for modelling (predictor variables)	Tumor size	✓	N/A	N/A	✓	✓	✓
	Tumor grade (G)	NS 1-9	Only G1 & G3	✓	✓ + nuclear grade	✓	✓
	LVI	N/A	N/A	N/A	✓	✓	N/A
	ER	0-300 H-score	2-tier <20% ≥20%	0–100%	2-tier <80% ≥80%	2-tier <1% ≥1%	2-tier <1% ≥1%
	PR	0-300 H-score	2-tier <1% ≥1%	0–100%	2-tier <80% ≥80%	2-tier <1% ≥1%	2-tier <1% ≥1%
	Ki-67 1–100%	✓	N/A	✓	N/A	N/A	N/A
	HER2 status	Pos Equiv Neg	Neg	Pos Neg	Neg	Neg	Neg
	Tumor histology	N/A	N/A	N/A	N/A	✓	✓
Oncotype DX scoreCut-off values (outcome, dependent variable)	Continuous 0-100	✓	N/A	N/A	N/A	N/A	N/A
	"Commercial" 0–17, 18–30, 31–100	✓	✓	N/A	✓	✓ 18–30 excluded	N/A
	TAILORx 0–10, 11–25, 26–100	N/A	✓	N/A	N/A	✓ 11–25 excluded	N/A
	TAILORx 0–25, 26–100	N/A	N/A	✓	N/A	N/A	✓
Statistics	Method	Multiple Linear Regression	2-step Discordance	Random forest model	Regression model→ Simplified scoring system	Binomial Logistic Regression	Binomial Logistic Regression
	Discrimination	R coeff ~0.60	No coeff for the model	No coeff for the model	C-index 0.854 (ODX>30 subset)	C-index 0.85; 0.88	C-index 0.81
	Calibration - % Accuracy (Model Predicted/ Observed)	54%–59%	<3% by 2-step discordance	52.5%	68%	80.4%	86.8%
On-line calculator		✓	N/A	✓	N/A	✓	✓

UPMC = University of Pittsburgh Medical Center "Magee equations", AAMC = "Anne Arundel Medical Center prediction tool", JH = John Hopkins "Breast cancer recurrence score estimator", MSKCC = Memorial Sloan Kettering Cancer Center "Inexpensive estimation tool", UTMC = University of Tennessee Medical Center "Breast cancer nomograms: Prediction for a low-risk and a high-risk Oncotype DX recurrence score", NS = Nottingham score, LVI = lymphovascular invasion, ✓ = used in the model.

patients with invasive breast carcinoma might not contribute to clinical management decisions. Some of these studies suggested that analysis of some of the clinicopathologic data available from pathology reports may select patients for which ODX test results were highly predictable and, therefore, the test could be safely and appropriately avoided [33,34].

We previously developed a "user friendly" nomogram/calculator [14] as a surrogate prediction model for the ODX score. That nomogram was based on clinicopathologic characteristics of ODX tested patients captured by the NCDB from 2010 to 2013. We did not use intermediate-risk ODXRS tested patients (11–25) in the creation of that nomogram, since at the time, guidelines for the role of adjuvant chemotherapy in this group of patients remained under investigation in an ongoing prospective TAILORx clinical trial.

For the development of the updated nomogram we used five clinicopathologic variables: age, tumor size, tumor grade, PR status and the histologic type of breast cancer. We tested their predictive value for forecasting the ODXRS while using the entire NCDB population of ODX tested patients. ODXRS cut-off values were based on the TAILORx clinical trial results [7]: 0–25 low-risk and 26–100 high-risk ODXRS, with intermediate-risk score patients (11–25) included in the low-risk score category. A multivariate logistic regression analysis of an updated nomogram revealed that all five clinicopathologic variables were significant predictors for a low-risk or a high-risk ODXRS test results with grade and PR status carrying the highest impact. Grade and PR status were also shown by Gage et al. [26,34] and Huang et al. [35] as significant predictors of ODXRS test results in different ODX prediction models.

The updated nomogram's c-index of 0.81 is considered a strong statistical model for both low-risk and high-risk ODXRS prediction outcomes [22,23]. The new online calculator https://utgsm.shinyapps.io/OncotypeDXCalculator/shows probabilities with 95% confidence intervals (CI) for a high-risk and low-risk ODX score. Prediction intervals based on 95% CI are very narrow (up to ±1–3% difference from the prediction values), attesting to the strength of our model built on the large number of patients.

Some benefit of adjuvant chemotherapy was identified in TAILORx trial [7] in patients younger than 50 years of age with ODXRS of 16–25. Therefore, we also evaluated our prediction model for ODXRS cut-off values 0–15 for a low-risk and 16–100 for a high-

risk ODXRS. Concordance index for these cut-off points was poor (c-index = 0.676) [23] so we did not use these cut-off points for building of the new nomogram. Inability of our model to discriminate between 0-15 and 16–100 cut-off points could be considered a weakness of our study. Nevertheless, when looking for the prediction for a high-risk ODXRS results (26–100), use of our updated nomograms is applicable for all ages. When looking for the prediction for a low-risk score, our updated nomograms are useful for patients older than 50. Therefore, overall utilization of our updated nomogram is still applicable for at least 75% of patients.

Development and utilization of genomic assays for breast cancer in clinical practice [6,7,36,37] has allowed us to identify ER+/HER2-negative/lymph node-negative breast cancer patients with early stage breast cancer for which the addition of chemotherapy to endocrine therapy is or is not beneficial. These assays have also allowed us to develop models for prediction of genomic assay results using readily available clinicopathologic variables, while avoiding the expense and delay of obtaining the actual assay.

Table 5 tabulates a summary of published studies that used sets of different clinicopathologic variables to predict for ODX test results. The comparison of published tools in predicting ODX test results with our own updated model would be the best measure to assess the additional value that our updated model offers. However, the table showed that a direct comparison of published tools with each other or with our own updated model is difficult for the following three reasons:

1) Each study had different sets of rules for choosing and scoring both predictor and outcome variables. For example, the studies used ER and PR scoring as the H-score (0–300) [30] or percentage of ER/PR positive cells (0–100%) [29], subsequently dichotomized to different percentage categories in two studies [26,38]. In our current study, ER and PR were reported as a 2-tier variable (<1%-negative and ≥1%-positive), since other ER and PR values were not available from the NCDB. Two studies used Ki-67 proliferation index (0–100%) [29,30], the variable not collected by the NCDB. Lack of consensus on reporting of Ki-67 proliferation index in breast cancer is considered an obstacle to its wide use in clinical practice, and is not a required element for prognosis of breast cancer [15]. This might hinder applicability of these ODX prediction tools in clinical practices in which Ki-67 proliferation index results are not available. For prediction of the outcome variable, ODX test results, cut off values chosen in building the prediction models were also different, as apparent in Table 5 further hampering accurate tool to tool comparison.
2) Statistical methods used to assess predictive ability of respective tools were also different, some being esoteric, such as the random forest model used by Kim et al. [29], further impeding accurate tool to tool comparison.
3) Reporting of the accuracy of the predictive models for predicting ODX test results was also different. Reported overall accuracy was between 54 and 68%, while our previously published nomogram [14] had the highest level of accuracy among published studies, at 80.4%. Overall accuracy of our current nomogram is 86.8%.

A recently published study by Roberson et al. addressed utilization of one of the ODX predictive models, Magee Equations (Table 5, Klein et al. [30]) for prediction of ODX assay results in clinical practice in Eastern Ontario [39]. They found that addition of Magee Equations in clinical practice did not decrease ordering of ODX testing. However, survey of oncologists participating in the study revealed that there was an increased physician comfort with systemic therapy decisions with availability of both Ki-67 proliferation index and Magee Equations. Our previously published nomogram [14] website visits reached 21,947 calculations performed in a period from 3/1/2017 to 10/26/2018. Through Orucevic's personal e-mail communications with oncologists from India and United Kingdom, it was learned that it is heavily utilized in their clinical practices. In an era of personalized medicine, ordering and utilization of ODX reflects a personalized clinician and patient decision-making process. Both the clinician and the patient need to "trust" that the ODX surrogate models that use clinicopathologic variables to predict ODX test results will give similar results as the ODX test itself. Therefore, the performance of the prediction tool in regards to calibration, discrimination, sensitivity, specificity, PPV and NPV needs to be easily understandable to both the clinician and the patient. One may wonder if addition of Magee Equations [30] in clinical practice in Eastern Ontario [39] did not decrease ordering of ODX testing because of clinician's uncertainty in regards to the accuracy of performance of Magee Equations. Or, there was no need for using a surrogate for ODX test when a "gold" standard ODX test offered by Genomic Health Inc., is available and affordable enough.

The strengths of our updated model of a nomogram/calculator for predicting ODX test results using clinicopathologic data from pathology reports are:

1) The model has been developed and validated from 84,339 patients' data collected by the National Cancer Data Base. Our model was built on the largest patient cohort ODX data to date, with data representing fair proportion of the United States NCDB patients.
2) The overall model accuracy for predicting ODX test results of 86.8% is the highest published to date (Table 5). In potentially clinically relevant probability range from 85% to 100%, accuracy of our prediction model rises to 92.7% (Table 4).
3) Our updated nomogram/calculator ODX prediction tool uses age, tumor size, grade, PR status and histologic type of breast cancer, the clinicopathologic variables that are required elements in breast cancer reporting and are easily obtained from any pathology report.
4) Our updated nomogram/calculator offers clear explanations of rigor included in the development, external validation, and performance. We incorporated all items relevant to the development and validation of our updated nomogram/calculator from the 22-item check list from the TRIPOD statement on "Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD)" [18], with transparent reporting of methodology and results. Therefore, our tool can easily be tested and validated in different clinical settings and other geographic locations. This also allows easy comparison with already published ODX prediction tools as well as with ones that will be developed in the future.
5) Since our updated nomogram now reflects ODX cut-off categories that are relevant for use of chemotherapy in ER+/HER-/lymph node-negative breast cancer patients based on the most recent published TAILORx clinical trial results [7], the updated nomogram is the most current with clinical practice recommendations in comparison to other published ODX prediction tools (Table 5).

Further research with refinement of genomic assays as well as more refined recording of clinicopathologic variables by large databases such as NCDB and Surveillance Epidemiology and End Results (SEER), will likely be beneficial in building more robust prediction models as surrogates for genomic assays based on clinicopathologic data. When these data become available, our updated nomogram model for prediction of ODX test results could be easily "re-trained".

Summary

An updated nomogram is now developed and validated based on the entire population of ODX tested patients (84,339) captured by the NCDB from 2010 to 2015. The updated nomogram's c-index of 0.81 is considered a strong statistical model for both low-risk and high-risk ODX recurrence score prediction outcomes. The model overall correctly assigns 86.8% of cases to the low- or high-risk ODX category. Furthermore, the overall accuracy of our prediction model rises to 92.7% at calculated probability range from 85% to 100%, which might be relevant for clinical use. This revised nomogram/calculator will continue serving as a surrogate for breast cancer patients for which Oncotype DX testing is not mandatory, affordable, or available.

Authors' declaration of interest

None.

Ethical approval

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

For our study, formal consent is not required. The NCDB de-identifies data that include the names of patients and institutions. Because of this, criteria of 45 CFR 46.102 d research are met, and Institutional Review Board approval is not required.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial or not-for-profit sectors.

Data availability statement

The data that support the findings of this study are available from the National Cancer Data Base, but restrictions apply to the availability of these data. These data were used under permission for the current study, and so are not publicly available. Data are however available from the authors upon reasonable request and with permission of the National Cancer Data Base.

Authors' contributions

Conception and design: Amila Orucevic, M.D., Ph.D.

Collection and assembly of data: Amila Orucevic, M.D., Ph.D., Megan King*, BS, Robert E. Heidel, Ph.D., Alison P. McNabb, MS.

*Note: Ms McNeil legally changed her last name now to King.

Data analysis and interpretation: All authors

Manuscript writing: All authors.

Final approval of manuscript: All authors.

Acknowledgement

Authors would like to thank Mr. Ryan King, BS in Mechanical Engineering, MS in Advanced Analytics, Analytics Professional/Sr. Statistician at General Motors, for his expertise and advice during development of the working version of updated calculator using R and Shiny application.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.breast.2019.05.006.

References

[1] Paik S, Shak S, Tang G, Kim C, Baker J, Cronin M, et al. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. N Engl J Med 2004;351:2817–26.
[2] Paik S, Tang G, Shak S, Kim C, Baker J, Kim W, et al. Gene expression and benefit of chemotherapy in women with node-negative, estrogen receptor-positive breast cancer. J Clin Oncol 2006;24:3726–34.
[3] Version 2 NCCN clinical practice guidelines in oncology (NCCN guidelines) Breast cancer. 2018. October 5, 2018, https://www.nccn.org/professionals/physician_gls/pdf/breast.pdf. [Accessed 8 October 2018].
[4] Harris LN, Ismaila N, McShane LM, Andre F, Collyar DE, Gonzalez-Angulo AM, et al. Use of biomarkers to guide decisions on adjuvant systemic therapy for women with early-stage invasive breast cancer: American society of clinical oncology clinical practice guideline. J Clin Oncol 2016;34(10):1134–50.
[5] AJCC cancer Staging Manual. 8th ed. Chicago, IL: Springer; 2016.
[6] Sparano JA, Gray RJ, Makower DF, Pritchard KI, Albain KS, Hayes DF, et al. Prospective validation of a 21-gene expression assay in breast cancer. N Engl J Med 2015;373:2005–14.
[7] Sparano JA, Gray RJ, Makower DF, Pritchard KI, Albain KS, Hayes DF, et al. Adjuvant chemotherapy guided by a 21-gene expression assay in breast cancer. N Engl J Med 2018;379:111–21.
[8] Orucevic A, Heidel RE, Bell JL. Utilization and impact of 21-gene recurrence score assay for breast cancer in clinical practice across the United States: Lessons learned from the 2010 to 2012 National Cancer Data Base analysis. Breast Canc Res Treat 2016;157:427–35.
[9] Albanell J, Svedman C, Gligorov J, Holt SD, Bertelli G, Blohmer JU, et al. Pooled Analysis of Prospective European Studies Assessing the Impact of Using the 21-Gene Recurrence Score Assay on Clinical Decision Making in Women With Oestrogen Receptor-Positive, Human Epidermal Growth Factor Receptor 2-Negative Early-Stage Breast Cancervol. 66. Oxford, England: European Journal of Cancer; 2016. p. 104–13.
[10] Dinan MA, Mi X, Reed SD, Hirsch BR, Lyman GH, Curtis LH. Initial trends in the use of the 21-gene recurrence score assay for patients with breast cancer in the medicare population, 2005-2009. JAMA Oncology 2015;1:158–66.
[11] Guth AA, Fineberg S, Fei K, Franco R, Bickell NA. Utilization of oncotype DX in an inner city population: Race or place? Int J Breast Cancer 2013;2013:653805.
[12] Roberts MC, Weinberger M, Dusetzina SB, Dinan MA, Reeder-Hayes KE, Carey LA, et al. Racial variation in the uptake of oncotype DX testing for early-stage breast cancer. J Clin Oncol 2016;34:130–8.
[13] Lund MJ, Mosunjac M, Davis KM, Gabram-Mendola S, Rizzo M, Bumpers HL, et al. 21-Gene recurrence scores: Racial differences in testing, scores, treatment, and outcome. Cancer 2012;118:788–96.
[14] Orucevic A, Bell JL, McNabb AP, Heidel RE. Oncotype DX breast cancer recurrence score can be predicted with a novel nomogram using clinico-pathologic data. Breast Canc Res Treat 2017;163:51–61.
[15] College of American Pathologists. Protocol for the Examination of Specimens From Patients With Invasive carcinoma of the Breast. Version: Invasive Breast 4.1.0.0. 2018.
[16] Iasonos A, Schrag D, Raj GV, Panageas KS. How to build and interpret a nomogram for cancer prognosis. J Clin Oncol 2008;26:1364–70.
[17] About National Cancer Data Base. https://www.facs.org/quality-programs/cancer/ncdb/about: American College of Surgeons.
[18] Moons KG, Altman DG, Reitsma JB, Ioannidis JP, Macaskill P, Steyerberg EW, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): Explanation and elaboration. Ann Intern Med 2015;162:W1–73.
[19] Harrell FEJ. RMS: Regression Modeling Strategies. R package version 3.5.1, http://CRAN.R-project.org/package=rms. 2018.
[20] Team RC. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing;https://www.R-project.org/2018.
[21] Harrell F. Regression Modeling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis. 2nd ed. New York: Springer; 2015.
[22] Alba AC, Agoritsas T, Walsh M, Hanna S, Iorio A, Devereaux PJ, et al. Discrimination and calibration of clinical prediction models: Users' guides to the medical literature. J Am Med Assoc 2017;318:1377–84.
[23] C-Statistic. Definition, Examples, Weighting and Significance. https://www.statisticshowto.datasciencecentral.com/c-statistic/2019.
[24] Breastcancer.org. Oncotype DX Test.
[25] Allison KH, Kandalaft PL, Sitlani CM, Dintzis SM, Gown AM. Routine pathologic parameters can predict Oncotype DX recurrence scores in subsets of ER positive patients: Who does not always need testing? Breast Canc Res Treat 2012;131:413–24.
[26] Gage MM, Rosman M, Mylander WC, Giblin E, Kim HS, Cope L, et al. A validated model for identifying patients unlikely to benefit from the 21-

gene recurrence score assay. Clin Breast Canc 2015;15:467–72.
[27] Geradts J, Bean SM, Bentley RC, Barry WT. The oncotype DX recurrence score is correlated with a composite index including routinely reported pathobiologic features. Cancer Invest 2010;28:969–77.
[28] Ingoldsby H, Webber M, Wall D, Scarrott C, Newell J, Callagy G. Prediction of Oncotype DX and TAILORx risk categories using histopathological and immunohistochemical markers by classification and regression tree (CART) analysis. Breast 2013;22:879–86.
[29] Kim HS, Umbricht CB, Illei PB, Cimino-Mathews A, Cho S, Chowdhury N, et al. Optimizing the use of gene expression profiling in early-stage breast cancer. J Clin Oncol 2016;34:4390–7.
[30] Klein ME, Dabbs DJ, Shuai Y, Brufsky AM, Jankowitz R, Puhalla SL, et al. Prediction of the Oncotype DX recurrence score: Use of pathology-generated equations derived by linear regression analysis. Mod Pathol 2013;26:658–64.
[31] Mattes MD, Mann JM, Ashamalla H, Tejwani A. Routine histopathologic characteristics can predict oncotype DX(TM) recurrence score in subsets of breast cancer patients. Cancer Invest 2013;31:604–6.
[32] Turner BM, Skinner KA, Tang P, Jackson MC, Soukiazian N, Shayne M, et al. Use of modified Magee equations and histologic criteria to predict the Oncotype DX recurrence score. Mod Pathol 2015;28:921–31.
[33] Dabbs DJ, Clark BZ, Serdy K, Onisko A, Brufsky AM, Smalley S, et al. Pathologist's health-care value in the triage of Oncotype DX((R)) testing: A value-based pathology study of tumour biology with outcomes. Histopathology 2018;73:692–700.
[34] Gage MM, Mylander WC, Rosman M, Fujii T, Le Du F, Raghavendra A, et al. Combined pathologic-genomic algorithm for early-stage breast cancer improves cost-effective use of the 21-gene recurrence score assay. Ann Oncol 2018;29:1280–5.
[35] Huang JL, Kizy S, Marmor S, Altman A, Blaes A, Beckwith H, et al. Tumor Grade and Progesterone Receptor Status Predict 21-Gene Recurrence Score in Early Stage Invasive Breast Carcinoma. Breast Cancer Research and Treatment; 2018.
[36] Cardoso F, van't Veer LJ, Bogaerts J, Slaets L, Viale G, Delaloge S, et al. 70-Gene signature as an aid to treatment decisions in early-stage breast cancer. N Engl J Med 2016;375:717–29.
[37] Sestak I, Buus R, Cuzick J, Dubsky P, Kronenwett R, Denkert C, et al. Comparison of the performance of 6 prognostic signatures for estrogen receptor-positive breast cancer: A secondary analysis of a randomized clinical trial. JAMA Oncology 2018;4:545–53.
[38] Eaton AA, Pesce CE, Murphy JO, Stempel MM, Patil SM, Brogi E, et al. Estimating the OncotypeDX score: Validation of an inexpensive estimation tool. Breast Canc Res Treat 2017;161:435–41.
[39] Robertson SJ, Ibrahim MFK, Stober C, Hilton J, Kos Z, Mazzarello S, et al. Does integration of Magee equations into routine clinical practice affect whether oncologists order the Oncotype DX test? A prospective randomized trial. J Eval Clin Pract 2019;25:196–204.