Annals of
SURGICAL ONCOLOGY
OFFICIAL JOURNAL OF THE SOCIETY OF SURGICAL ONCOLOGY

ORIGINAL ARTICLE – BREAST ONCOLOGY

Performance and Clinical Utility of Models Predicting Eradication of Nodal Disease in Patients with Clinically Node-Positive Breast Cancer Treated with Neoadjuvant Chemotherapy by Tumor Biology

John Davis Jr., MD[1], Tanya L. Hoskin, MS[2], Courtney N. Day, BS[2], Mark Wickre, MD[3],
Mara A. Piltin, DO[1], Abigail S. Caudle, MD[4], and Judy C. Boughey, MD[1]

[1]Department of Surgery, Mayo Clinic, Rochester, MN; [2]Department of Health Science Research and Clinical Statistics, Mayo Clinic, Rochester, MN; [3]Department of Radiology, Mayo Clinic, Rochester, MN; [4]Department of Breast Surgical Oncology, University of Texas MD Anderson Cancer Center, Houston, TX

ABSTRACT

Introduction. Prediction models are useful to guide decision making. Our goal was to compare three published nomograms predicting axillary response to neoadjuvant chemotherapy (NAC), clinically node-positive breast cancer.

Methods. Patients with cT1–T4, cN1–N3 breast cancer treated with NAC and surgery from 2008 to 2019 were reviewed. The predicted probability of pathologic node-negative (ypN0) status was estimated for each nomogram. Area under the curve (AUC) was compared across models, overall and by biologic subtype.

Results. Of 581 patients, 253 (43.5%) were ypN0. ypN0 status varied by subtype: 23.9% for estrogen receptor-positive (ER+)/human epidermal growth factor receptor 2-negative (HER2−), 68.9% for HER2-positive (HER2+), and 47.2% for ER-negative (ER−)/HER2−. The three nomograms had similar AUC values (0.761–0.769; $p = 0.80$). The Mayo model-predicted probability was significantly lower ($p < 0.001$) than the observed probability of ypN0 status, while the MD Anderson Cancer

Center (MDACC) 1- and 2-predicted probabilities were similar to the observed probability. At a predicted probability threshold of 50%, the Mayo model had the highest sensitivity (89.6%) for detecting ypN+ patients compared with MDACC models 1 and 2 (76.5%; $p < 0.001$). However, both MDACC models had higher specificity in identifying ypN0 status among HER2+ (81.7%) and ER−/HER2− (75.9–77.6%) patients compared with the Mayo model (59.5% and 43.1%; each $p < 0.001$). None of the models identified the ER+/HER2− patients with ypN0 status well at the $\geq 50\%$ threshold (specificity 0–9.4%).

Conclusion. All three models predicting nodal response to NAC performed well overall with respect to discrimination, but differed with respect to calibration and performance at a 50% probability threshold. However, none of the models performed well at the 50% threshold for ER+/HER2− patients.

John Davis Jr and Tanya L. Hoskin are joint first authors.

As management of breast cancer advances, incorporating tumor biology in decision making is paramount. Tumor biology, response to neoadjuvant chemotherapy (NAC), and nodal status are several key factors that help decide the patient's optimal therapy. Various clinical tools and nomograms are available online that incorporate specific tumor factors to help with treatment decisions.[1–6] The use of NAC in operable breast cancer continues to increase, especially in triple-negative breast cancer (TNBC) and human epidermal growth factor receptor 2 (HER2)-positive (HER2+) disease.[7] NAC allows for in vivo assessment of

treatment response and can result in downstaging disease in the breast and axilla, and de-escalation of breast and axillary surgery.[7–10]

Several prospective studies have led to a change in the way surgeons manage the axilla in breast cancer patients treated with NAC. The SENTINA trial, SN FNAC trial, and the American College of Surgeons Oncology Group (ACOSOG) Z1071 trial showed that sentinel lymph node (SLN) surgery can be performed for node-positive women with a good response to NAC, and has led to an increase in the use of SLN surgery for these patients.[11–13]

When considering SLN surgery after NAC for a patient with cN+ disease at diagnosis, the patients most likely to benefit from this approach are those most likely to have a complete nodal response. Preoperative models that predict the likelihood of the patient achieving a pathologic complete response (pCR) in the axilla after NAC are helpful to guide this decision making. Multiple models have been published predicting axillary pCR after NAC in varying cohorts.[2–6] Authors at the Mayo Clinic used the National Cancer Database (NCDB) from 2010 to 2012 to construct two models that showed good discrimination for predicting (pathologic node-negative) pN0 status following NAC in patients who were clinically node-negative (cN0) and clinically node-positive (cN+) at presentation.[9] The model for patients with cN+ disease had an area under the receiver operating characteristic curve (AUC) of 0.71 and was created using data from 9494 patients, of whom 3314 (34.9%) achieved a nodal pCR. In another study from MD Anderson Cancer Center (MDACC), 584 cT1–4, N1, M0 patients who underwent NAC from 2001 to 2013 were studied. Two predictive models (AUC 0.765 and 0.776) were created using their database. In their model, 37.2% of patients had nodal pCR.[3]

The goal of the current study was to compare three validated predictive models for nodal response in patients with cN+ disease treated with NAC to assess model generalizability to a different setting and to compare model performance and clinical utility, both overall and by biologic subtype, in our patient cohort.

METHODS

With Institutional Review Board approval, patients with cT1–T4, cN1–N3 breast cancer who underwent surgical treatment with axillary staging after NAC at Mayo Clinic Rochester from 2008 to 2019 were reviewed. All cN+ patients included were pathologically confirmed by percutaneous biopsy. Patients with distant metastasis or prior history of breast cancer were excluded. Data regarding patient demographics, tumor biology, and surgical procedure were obtained from the breast surgery registry, and

imaging studies were reviewed for number of abnormal nodes on axillary ultrasound. Patients were classified as nodal pCR (ypN0) if they had negative nodes or isolated tumor cells, and as ypN+ if they had micrometastasis or larger (> 0.2 mm) at surgery after NAC. The predicted probability of ypN0 status was calculated using logistic regression model coefficients from the three published models; these coefficients did not appear in the original MDACC publication but were provided by the study authors upon request. The Mayo predictive model included age, clinical tumor category (cT), clinical nodal status (cN), grade, histology, and tumor biologic subtype (estrogen receptor-positive [ER+]/HER2-negative [HER2−], ER+/HER2+, ER-negative [ER−]/HER2+, and ER−/HER2−). The MDACC has two predictive models. The MDACC 1 predictive model included clinical tumor category (cT), multifocal/multicentric disease, number of pretreatment sonographically suspicious nodes (< 4 vs. ≥ 4), nuclear grade, HER2±, ER± and progesterone receptor (PR)±. The MDACC 2 predictive model was similar to MDACC 1 except it did not include the number of suspicious nodes as a variable, but did include histology (lobular vs. mixed or ductal).

Statistical Analysis

To assess discrimination, the AUC was estimated and was reported with 95% confidence intervals for each model. AUCs were compared across models using the DeLong error method. Calibration was assessed by comparing the mean predicted probability from each model to the observed probability of ypN0 status using a one-sample t test and via loess calibration curves showing the relationship between the observed probability and the predicted probability on a continuous scale. To assess utility for clinical decision making, we classified each patient as having ≥ 50% or < 50% probability of ypN0 status by each model. This classification was then compared with actual pathologic node status to estimate specificity (i.e. probability that a ypN0 patient was predicted as having ≥ 50% probability of node-negative status) and sensitivity (i.e. probability a ypN+ patient was predicted as having < 50% probability of node-negative status). Using a probability of 50% likely represents a common rationale when reducing a continuous probability to a binary clinical choice—in effect dichotomizing, according to which of the two outcomes is most likely. In addition to the cut-point of 50%, we also identified the optimal cut-point for each model as the value that minimized the absolute difference between sensitivity and specificity, and reported performance metrics also at that cut-point. Sensitivity and specificity values were compared between models using McNemar's test for paired

proportions. *P*-values < 0.05 were considered statistically significant. Analysis was performed using SAS version 9.4 (SAS Institute Inc., Cary, NC, USA) and R version 3.6.1 (The R Foundation for Statistical Computing, Vienna, Austria).

RESULTS

Overall, 581 patients with cT1–4, N1–3 breast cancer treated with NAC and meeting the study inclusion criteria were identified. Median patient age was 50 years (range 24–86). Clinical tumor category distribution was 13.3% cT1, 49.7% cT2, 32.4% cT3, and 4.6% cT4. The majority of patients (519 [89.3%]) in the study had cN1 disease, and a minority (208 [35.8%]) were multicentric or multifocal in nature. Of the 581 patients, 156 (26.9%) had SLN only, 146 (25.1%) had SLN followed by completion axillary lymph node dissection (ALND), and 279 (48.0%) had ALND only. The median number of nodes removed in cases with SLN only was four. The nodal pCR (ypN0) rate was 43.5%, with 188 (32.4%) patients with ypN1 disease, 90 (15.5%) with ypN2 disease, and 50 (8.6%) with ypN3 disease (Table 1).

Comparing patients who achieved a nodal pCR (ypN0) and those with residual nodal disease (ypN+), there was no difference in clinical N category at presentation, or the proportion of patients with fewer than four abnormal lymph nodes on ultrasound (Table 1). Nodal pCR was higher with smaller tumors at presentation (cT1 and cT2). Higher grade tumors (more poorly differentiated) were more likely to achieve nodal pCR. Younger patients (median age 49 vs. 52 years; *p* = 0.006) had higher nodal pCR rates. Biological subtypes played a significant role in the nodal pCR rate, with HER2+ disease having the highest rate of nodal pCR (68.9%) and ER +/HER2− having the lowest rate (23.9%; *p* < 0.001) (Table 1).

Overall, the three predictive nomograms had very similar AUC values, ranging from 0.761 to 0.769 (*p* = 0.80), with good discrimination. Given the importance of tumor biologic subtype, we further assessed model performance separately by biologic subtype and found that the AUCs within the subtypes were similar across the three models (Table 2 and Fig. 1). Not surprisingly, the within-subtype AUC values were generally lower than for the overall cohort, ranging from 0.67 to 0.70 in ER +/HER2−, 0.64–0.66 in HER2+, and 0.66–0.68 in ER−/HER2−.

To assess model calibration, the mean predicted probability from each model was compared with the observed probability (Table 3). The predicted probability from the Mayo model (mean 34.9%) was significantly lower (*p* < 0.001) than the observed probability of ypN0 status in the cohort (43.5%), while MDACC 1- and 2-predicted probabilities (mean 42.3% and 41.7%, respectively) were similar to the observed probability. This poorer calibration for the Mayo model is also apparent in calibration curves (Fig. 2a), where the Mayo model estimated a lower probability of ypN0 status than the observed probability over much of the range, whereas the MDACC models showed more consistent concordance between predicted and observed probability. When examined based on biological subtypes, HER2+ patients have the highest predicted probabilities, followed by the ER−/HER− patients (Table 3, Figs. 2b–d), while the predicted probabilities of ypN0 status were markedly lower for ER+/HER2− patients in all models.

Considering specificity, at a cut-point of ≥ 50% predicted probability of ypN0 status, MDACC 1 and 2 each identified 157/253 (62.1%) ypN0 patients as having ≥ 50% probability of negative nodes, whereas the Mayo model identified only 103/253 (40.7%; *p* < 0.001). By subtype, MDACC 1 predicted probability ≥ 50% identified 107/131 (81.7%) HER2+ patients and 45/58 (77.6%) ER−/HER2− patients with ypN0 status. MDACC 2 performed similarly, while the Mayo model identified only 78/131 (59.5%) and 25/58 (43.1%) of patients with nodal pCR, respectively (*p* < 0.001). For the ER+/HER2− patients with ypN0 status, none of the three models identified them well at the ≥ 50% threshold (specificity 0–9.4%). Considering sensitivity, at the threshold of < 50% predicted probability of ypN0 status, the Mayo model identified a higher percentage of ypN+ patients (294/328, 89.6%) compared with the MDACC 1 and 2 models (251/328, 76.5%; *p* < 0.001) (Table 2).

Using optimal cut-points that maximize sensitivity and specificity instead of the arbitrary 50% threshold, the optimal thresholds varied by biologic subtype (Table 3) and were highest in HER2+, followed by ER−/HER2−, and substantially lower (21–22%) in ER+/HER2−. The reason for this is clear when examining the distribution of predicted probabilities by subtype (Fig. 2b–d) split by ypN status. The distribution of predicted probabilities for a given subtype is strongly related to the prevalence of ypN0 status in that subtype; thus, almost no ER+/HER2− patients have ≥ 50% probability of nodal pCR when considered among the cohort as a whole, making this cut-point not useful for discriminating between ypN0 and ypN+ patients in this subtype. However, applying a lower threshold, such as ≥ 22% predicted probability, MDACC 1 achieved useful results for identifying patients with a higher likelihood of ypN0 status within the ER+/HER2− subtype, with sensitivity and specificity of 65.2% and 65.6%, respectively (Table 3).

TABLE 1 Demographic and clinical characteristics of the cohort including 581 patients treated with NAC for clinically node-positive breast cancer, overall and by pathologic node status

	Total (N = 581)	ypN0 (n = 253)	ypN+ (n = 328)	p-value
Age at surgery [median (range)]	50 (24–86)	49 (26–86)	52 (24–82)	0.006[a]
Age at surgery, years (Group)				0.03[b]
< 40	91 (15.7%)	47 (18.6%)	44 (13.4%)	
40–49	183 (31.5%)	85 (33.6%)	98 (29.9%)	
50–59	168 (28.9%)	75 (29.6%)	93 (28.4%)	
60–69	106 (18.2%)	32 (12.6%)	74 (22.6%)	
70 +	33 (5.7%)	14 (5.5%)	19 (5.8%)	
Clinical T category				0.01[b]
cT1	77 (13.3%)	40 (15.8%)	37 (11.3%)	
cT2	289 (49.7%)	138 (54.5%)	151 (46.0%)	
cT3	188 (32.4%)	66 (26.1%)	122 (37.2%)	
cT4	27 (4.6%)	9 (3.6%)	18 (5.5%)	
Clinical N category				0.54[b]
cN1	519 (89.3%)	230 (90.9%)	289 (88.1%)	
cN2	20 (3.4%)	7 (2.8%)	13 (4.0%)	
cN3	42 (7.2%)	16 (6.3%)	26 (7.9%)	
Number of sonographically suspicious nodes				0.58[b]
< 4	402 (69.2%)	172 (68.0%)	230 (70.1%)	
≥ 4	179 (30.8%)	81 (32.0%)	98 (29.9%)	
Multicentric and/or multifocal				0.004[b]
No	373 (64.2%)	179 (70.8%)	194 (59.1%)	
Yes	208 (35.8%)	74 (29.2%)	134 (40.9%)	
Grade				< 0.001[b]
I (well differentiated)	28 (4.8%)	4 (1.6%)	24 (7.3%)	
II (moderately differentiated)	251 (43.2%)	78 (30.8%)	173 (52.7%)	
III (poorly differentiated)	302 (52.0%)	171 (67.6%)	131 (39.9%)	
Histology				0.007[c]
IDC	498 (85.7%)	226 (89.3%)	272 (82.9%)	
ILC	37 (6.4%)	7 (2.8%)	30 (9.1%)	
IMC	43 (7.4%)	18 (7.1%)	25 (7.6%)	
Other	3 (0.5%)	2 (0.8%)	1 (0.3%)	
ER				< 0.001[b]
Negative	194 (33.4%)	117 (46.2%)	77 (23.5%)	
Positive	387 (66.6%)	136 (53.8%)	251 (76.5%)	
PR				< 0.001[b]
Negative	240 (41.3%)	138 (54.5%)	102 (31.1%)	
Positive	341 (58.7%)	115 (45.5%)	226 (68.9%)	
HER2				< 0.001[b]
Negative	391 (67.3%)	122 (48.2%)	269 (82.0%)	
Positive	190 (32.7%)	131 (51.8%)	59 (18.0%)	
Biologic subtype				< 0.001[b]
ER+/HER2−	268 (46.1%)	64 (25.3%)	204 (62.2%)	
ER−/HER2−	123 (21.2%)	58 (22.9%)	65 (19.8%)	
HER2+	190 (32.7%)	131 (51.8%)	59 (18.0%)	

TABLE 1 continued

	Total (N = 581)	ypN0 (n = 253)	ypN+ (n = 328)	p-value
Pathologic T category				< 0.001[b]
ypT0	150 (25.8%)	131 (51.8%)	19 (5.8%)	
ypTis	50 (8.6%)	38 (15.0%)	12 (3.7%)	
ypT1	176 (30.3%)	65 (25.7%)	111 (33.8%)	
ypT2	122 (21.0%)	15 (5.9%)	107 (32.6%)	
ypT3	72 (12.4%)	3 (1.2%)	69 (21.0%)	
ypT4	11 (1.9%)	1 (0.4%)	10 (3.0%)	
Pathologic N category				
ypN0	253 (43.5%)	253 (100.0%)	0	
ypN1	188 (32.4%)	0	188 (57.3%)	
ypN2	90 (15.5%)	0	90 (27.4%)	
ypN3	50 (8.6%)	0	50 (15.2%)	
Pathologic node status				
ypN0	253 (43.5%)			
ypN+	328 (56.5%)			

Data are expressed as n (%) unless otherwise specified

NAC neoadjuvant chemotherapy, *IDC* invasive ductal carcinoma, *ILC* invasive lobular carcinoma, *IMC* invasive mammary carcinoma, *ER* estrogen receptor, *PR* progesterone receptor, *HER2* human epidermal growth factor receptor 2

[a]Wilcoxon rank-sum test

[b]Chi square test

[c]Fisher's exact test

TABLE 2 Performance of three predictive models, overall and separately for each biologic subtype

|	AUC (95% CI)	Performance at the nominal threshold of ≥ 50% predicted probability ypN0		Performance at the optimal[a] threshold for predicted probability ypN0		
		Sensitivity[b]	Specificity[c]	Threshold	Sensitivity[b]	Specificity[c]
Entire cohort						
Mayo	0.761 (0.722–0.800)	89.6	40.7	38	69.2	68.8
MDACC 1	0.768 (0.730–0.807)	76.5	62.1	42	68.9	70.4
MDACC 2	0.769 (0.731–0.807)	76.5	62.1	42	69.5	70.4
ER+/HER2−						
Mayo	0.669 (0.592–0.746)	100	0	21	58.3	54.7
MDACC 1	0.702 (0.632–0.771)	96.6	7.8	22	65.2	65.6
MDACC 2	0.704 (0.636–0.772)	96.1	9.4	21	63.2	65.6
HER2+						
Mayo	0.655 (0.572–0.739)	66.1	59.5	49	64.4	62.6
MDACC 1	0.636 (0.549–0.723)	39.0	81.7	66	54.2	64.9
MDACC 2	0.652 (0.566–0.738)	37.3	81.7	64	59.3	64.1
ER−/HER2−						
Mayo	0.659 (0.563–0.755)	78.5	43.1	47	58.5	60.3
MDACC 1	0.678 (0.584–0.772)	47.7	77.6	56	61.5	67.2
MDACC 2	0.664 (0.568–0.759)	50.8	75.9	57	61.5	67.2

AUC area under the receiver operating characteristic curve, *CI* confidence interval, *MDACC* MD Anderson Cancer Center, *ER* estrogen receptor, *HER2* human epidermal growth factor receptor 2

[a]Optimal threshold selected as the value with the minimum absolute difference between sensitivity and specificity

[b]Sensitivity defined as the percentage of ypN+ patients with predicted probability of negative nodes less than the threshold

[c]Specificity defined as the percentage of ypN0 patients with predicted probability of negative nodes greater than or equal to the threshold

FIG. 1 Receiver operating characteristic curves for the prediction of ypN0 status for three clinical prediction models, assessed for the entire cohort of 581 patients treated with NAC, and also separately by biologic subtype. *NAC* neoadjuvant chemotherapy, *MDACC* MD Anderson Cancer Center, *ER+* estrogen receptor-positive, *ER−* estrogen receptor-negative, *HER2* human epidermal growth factor receptor 2, *HER2−* HER2-negative, *HER2+* HER2-positive, *TNBC* triple-negative breast cancer
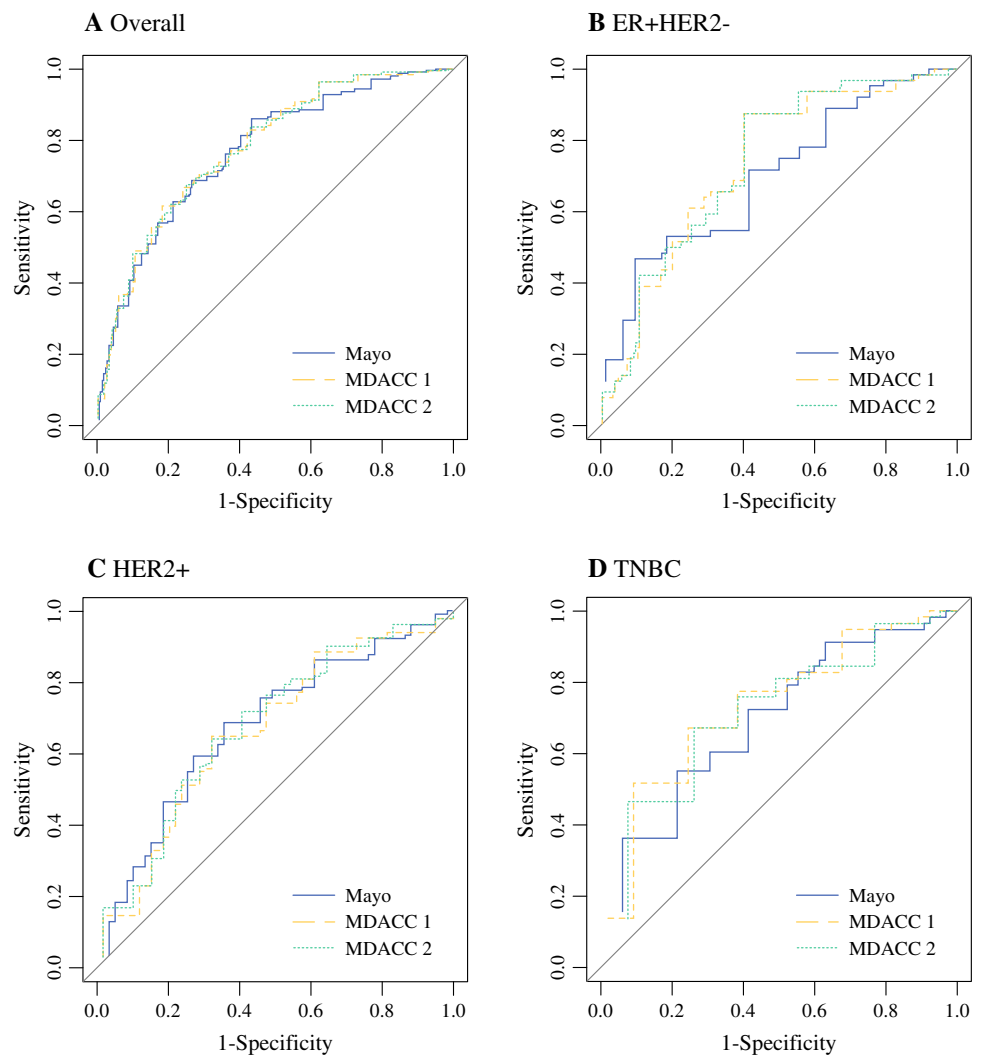
DISCUSSION

Predictive models that estimate the likelihood of converting from cN+ prior to NAC to ypN0 after NAC may be useful for clinical decision making. In this study, we compared the performance of three previously validated predictive models using readily available clinical variables. All three models performed well overall with respect to discrimination. Calibration was superior for the MDACC models; however, further differences were identified when assessing performance at the cut-point of 50% predicted probability. At this cut-point, the Mayo model had better sensitivity for correctly classifying patients with ypN+, while the MDACC models generally had better specificity for identifying patients with ypN0; however, all models had poor specificity (< 10%) in the ER+/HER2− subset at this threshold.

The MDACC models performed better than the Mayo model with respect to calibration, likely because the prevalence of ypN0 status in the MDACC model development sample (37.2%) was closer to the prevalence in the sample assessed here (43.5%), whereas the NCDB sample from which the Mayo model was built had a lower prevalence of nodal pCR (34.9%). Correspondingly, the predicted probabilities of ypN0 status were generally higher in the MDACC 1 and 2 models than the Mayo model. This emphasizes the importance of validating the generalizability and transferability of a predictive model developed in one setting to another setting.[14] The MDACC model may have performed better in the Mayo patient cohort because of the similarity of their clinical settings, since both are academic and tertiary referral centers. However, the Mayo model was developed in the NCDB, which includes data from more than 1500 Commission on Cancer facilities in diverse community and academic

TABLE 3 Comparison of the mean model-predicted probability of ypN0 status with the observed percentage of ypN0 to assess model calibration, overall and by biologic subtype

	Mean predicted probability of ypN0 (%)	Observed ypN0 (%)	p-value
Overall			
Mayo	34.9	43.5	< 0.001
MDACC 1	42.3	43.5	0.22
MDACC2	41.7	43.5	0.07
ER+/HER2−			
Mayo	19.1	23.9	< 0.001
MDACC 1	21.8	23.9	0.003
MDACC2	21.3	23.9	< 0.001
HER2+			
Mayo	50.2	68.9	< 0.001
MDACC 1	65.6	68.9	0.009
MDACC2	64.8	68.9	0.001
ER−/HER2−			
Mayo	45.4	47.2	0.008
MDACC 1	50.9	47.2	0.002
MDACC2	50.4	47.2	0.006

MDACC MD Anderson Cancer Center, *ER* estrogen receptor, *HER2* human epidermal growth factor receptor 2

FIG. 2 **a** Loess calibration curves showing the relationship between the model predicted probability of ypN0 status and the observed probability for the cohort of 581 patients treated with NAC. **b–d** Side-by-side boxplots comparing the predicted probability distribution between ypN0 and ypN+ patients separately, by subtype for the **b** Mayo model, **c** MDACC 1 model, and **d** MDACC 2 model. *NAC* neoadjuvant chemotherapy, *MDACC* MD Anderson Cancer Center, *ER+* estrogen receptor-positive, *ER−* estrogen receptor-negative, *HER2* human epidermal growth factor receptor 2, *HER2−* HER2-negative, *HER2+* HER2-positive, *TNBC* triple-negative breast cancer
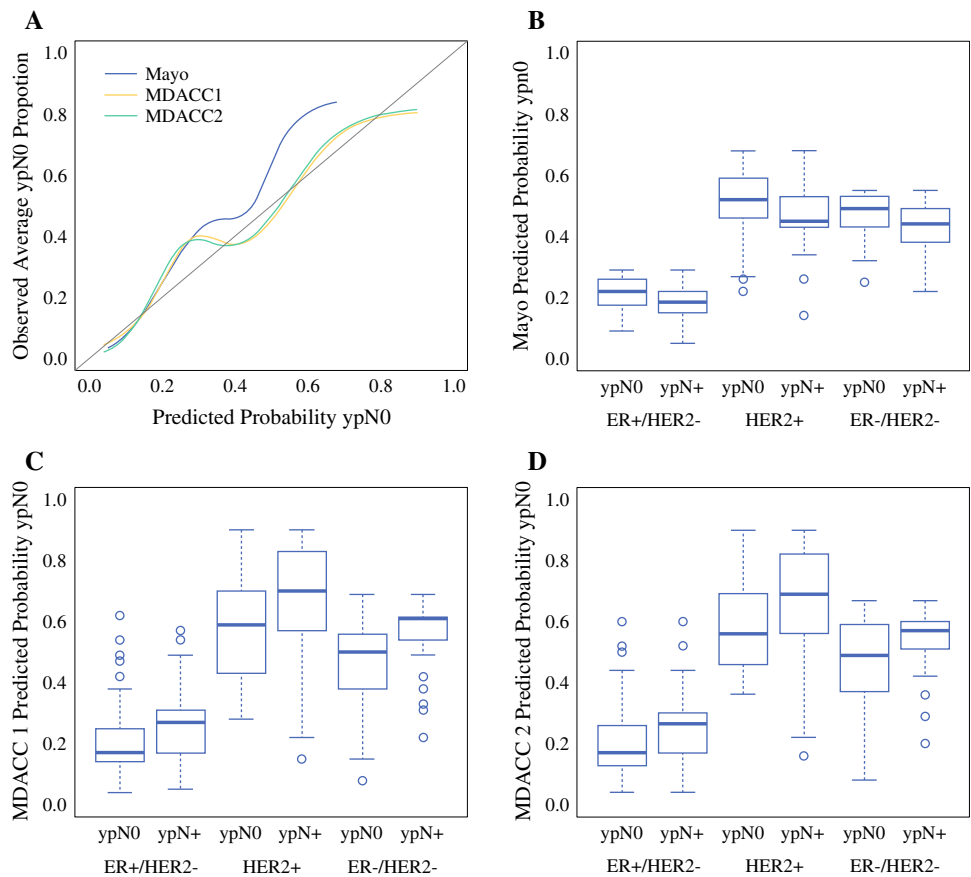
settings nationwide; thus, it is possible that the Mayo model is more relevant for the general population and that its calibration would be better in other settings.

For the MDACC 1 and 2 models, a key difference between the two models was the inclusion of the number of suspicious nodes on baseline axillary ultrasound. We found

very similar results for these two models and little benefit from a statistical perspective to adding the ultrasound findings. This is important, especially in some institutions where ultrasound data are not readily available or documented precisely, and a clinical prediction model that does not require this information could be useful.

The percentage of patients with ypN0 status varied markedly by subtype as expected.[11,][15–19] Likely due to availability of targeted therapy, HER2+ patients had the highest response, with 68.9% ypN0. ER−/HER2− and ER+/HER− patients had ypN0 rates of 47.2% and 23.9%, respectively. These data were mirrored in the predicted probability of ypN0 status for each subtype. We saw that HER2+ had the highest predicted probabilities, followed by ER−/HER2− and, lastly, ER+/HER−. This is important because it explains why a probability cut-point of 50% uniformly applied did not perform equally well in all subtypes and identified very few (< 10%) of the ypN0 patients in the ER+/HER2− subtype who might be candidates for SLN surgery. A lower probability of nodal complete response is inherent in the ER+/HER2− subtype, making 50% a poor threshold for decision making within this subtype since ≤ 5% of ER+/HER2− patients had ≥ 50% predicted probability of ypN0 status by any model. Thus, using optimal, subtype-specific cut-points provided discrimination between ypN0 and ypN+ patients in ER+/HER2− patients similar to that found in other subtypes. Because tumor biology was such a strong discriminatory feature, the AUCs for the models applied to the entire cohort were also higher than the within-subtype AUCs, which were generally only fair and < 0.70. This represents a possible opportunity to improve performance with the development of subtype-specific prediction models.

The Mayo model performed better when assigning a probability < 50% of ypN0 status to ypN+ patients with sensitivity of 89.6% compared with both MDACC models at 76.5%; higher sensitivity means the model would appropriately select a higher percentage of ypN+ patients for ALND. The MDACC 1 and 2 models are more specific than the Mayo model and were thus more likely to identify ypN0 patients as having ≥ 50% predicted probability of nodal pCR, resulting in an accurate choice to perform SLN surgery and spare the patient ALND. Considering the cost–benefit ratio in this situation, where the choice is to perform SLN surgery or not, one could argue a model with higher specificity that correctly identifies more of the ypN0 patients (to consider SLN surgery) is superior to a model that identifies fewer of the ypN0 patients. Alternatively, choosing the more sensitive model better identifies patients who are likely to remain ypN+, and therefore could be considered for ALND to limit operating time and mitigate the false negative rate.

The clinical use of these predictive models is in their utility in counseling patients about the likelihood of complete nodal response following NAC. This gives patients and physicians the ability to have a frank discussion in the preoperative setting regarding their specific surgical management. For patients at high likelihood of complete nodal response, they can be considered for less extensive axillary surgery, such as SLN surgery with or without targeting the clipped node, or the MARI procedure (marking the axillary lymph node with radioactive iodine seeds).[20]

Limitations of this study include the use of prospectively collected data evaluated in a retrospective fashion. Several other prediction models have also been published to predict the likelihood of node-negative disease after NAC for patients with node-positive disease.[1–6] Our study was limited to those models using clinical variables standardly available in the clinical record, and did not compare all available models in the literature. All three models that we evaluated have unique features that may be clinically obtained and universally applicable to patients in all types of medical centers. We feel that the benefit of increased clinical utility outweighs the limitations when using predictive nomograms as it relates to axillary management after NAC.

CONCLUSION

The three models to predict nodal response to NAC all performed well overall with respect to discrimination, demonstrating their clinical utility to select patients likely to benefit from SLN surgery, especially patients with ER−/HER2− or HER2+ disease. At a 50% threshold, MDACC models 1 and 2 had better specificity than the Mayo model; however, the Mayo model had better sensitivity. None of the models performed well for ER+/HER2− patients at the 50% threshold.

REFERENCES

1. Barron AU, Hoskin TL, Boughey JC. Predicting non-sentinel lymph node metastases in patients with a positive sentinel lymph node after neoadjuvant chemotherapy. *Ann Surg Oncol.* 2018;25(10):2867–74.
2. Kantor O, Sipsy LM, Yao K, James TA. A predictive model for axillary node pathologic complete response after neoadjuvant chemotherapy for breast cancer. *Ann Surg Oncol.* 2018;25(5):1304–11.
3. Vila J, Mittendorf EA, Farante G, et al. Nomograms for predicting axillary response to neoadjuvant chemotherapy in clinically node-positive patients with breast cancer. *Ann Surg Oncol.* 2016;23(11):3501–9.

4. Schipper RJ, Moossdorff M, Nelemans PJ, et al. A model to predict pathologic complete response of axillary lymph nodes to neoadjuvant chemo(immuno)therapy in patients with clinically node-positive breast cancer. *Clin Breast Cancer.* 2014;14(5):315–22.

5. Choi HJ, Ryu JM, Kim I, et al. Nomogram for accurate prediction of breast and axillary pathologic response after neoadjuvant chemotherapy in node positive patients with breast cancer. *Ann Surg Treat Res.* 2019;96(4):169–76.

6. Kim JY, Park HS, Kim S, Ryu J, Park S, Kim SI. Prognostic nomogram for prediction of axillary pathologic complete response after neoadjuvant chemotherapy in cytologically proven node-positive breast cancer. *Medicine (Baltimore).* 2015;94(43):e1720.

7. Murphy BL, Day CN, Hoskin TL, Habermann EB, Boughey JC. Neoadjuvant chemotherapy use in breast cancer is greatest in excellent responders: triple-negative and HER2 + subtypes. *Ann Surg Oncol.* 2018;25(8):2241–8.

8. King TA, Morrow M. Surgical issues in patients with breast cancer receiving neoadjuvant chemotherapy. *Nat Rev Clin Oncol.* 2015;12(6):335–43.

9. Untch M, Konecny GE, Paepke S, von Minckwitz G. Current and future role of neoadjuvant therapy for breast cancer. *Breast.* 2014;23(5):526–37.

10. Murphy BL, Hoskin TL, Heins CDN, Habermann EB, Boughey JC. Preoperative prediction of node-negative disease after neoadjuvant chemotherapy in patients presenting with node-negative or node-positive breast cancer. *Ann Surg Oncol.* 2017;24(9):2518–25.

11. Boughey JC, Suman VJ, Mittendorf EA, et al. Sentinel lymph node surgery after neoadjuvant chemotherapy in patients with node-positive breast cancer: the ACOSOG Z1071 (Alliance) clinical trial. *JAMA.* 2013;310(14):1455–61.

12. Kuehn T, Bauerfeind I, Fehm T, et al. Sentinel-lymph-node biopsy in patients with breast cancer before and after neoadjuvant chemotherapy (SENTINA): a prospective, multicentre cohort study. *Lancet Oncol.* 2013;14(7):609–18.

13. Boileau JF, Poirier B, Basik M, et al. Sentinel node biopsy after neoadjuvant chemotherapy in biopsy-proven node-positive breast cancer: the SN FNAC study. *J Clin Oncol.* 2015;33(3):258–64.

14. Steyerberg EW, Harrell FE Jr. Prediction models need appropriate internal, internal–external, and external validation. *J Clin Epidemiol.* 2016;69:245–47.

15. Kuerer HM, Sahin AA, Hunt KK, et al. Incidence and impact of documented eradication of breast cancer axillary lymph node metastases before surgery in patients treated with neoadjuvant chemotherapy. *Ann Surg.* 1999;230(1):72–8.

16. Buzdar AU, Ibrahim NK, Francis D, et al. Significantly higher pathologic complete remission rate after neoadjuvant therapy with trastuzumab, paclitaxel, and epirubicin chemotherapy: results of a randomized trial in human epidermal growth factor receptor 2-positive operable breast cancer. *J Clin Oncol.* 2005;23(16):3676–85.

17. Dominici LS, Negron Gonzalez VM, Buzdar AU, et al. Cytologically proven axillary lymph node metastases are eradicated in patients receiving preoperative chemotherapy with concurrent trastuzumab for HER2-positive breast cancer. *Cancer.* 2010;116(12):2884–89.

18. Hennessy BT, Hortobagyi GN, Rouzier R, et al. Outcome after pathologic complete eradication of cytologically proven breast cancer axillary node metastases following primary chemotherapy. *J Clin Oncol.* 2005;23:9304–11.

19. Boughey JC, McCall LM, Ballman KV, et al. 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 (**discussion 614–6**).

20. Donker M, Straver ME, Wesseling J, et al. Marking axillary lymph nodes with radioactive iodine seeds for axillary staging after neoadjuvant systemic treatment in breast cancer patients: the MARI procedure. *Ann Surg.* 2015;261(2):378–82.