Pathobiologic Stratification of Oncotype DX Recurrence Scores and Comparative Validation of 3 Surrogate Models

Anas Mohamed, MD; Aisha Kousar, MD; Jan Wong, MD; Nasreen Vohra, MD; Mahvish Muzaffar, MD; Joseph Geradts, MD

• ***Context.***—**The Oncotype DX Recurrence Score (RS) predicts recurrence and chemotherapy benefit in early-stage estrogen receptor–positive breast cancer patients. Cost and unavailability are 2 major disadvantages of the assay. Multiple models have been developed to predict the RS.**

Objective.—**To predict RS based on histopathologic and biomarker features, and to measure concordance and correlation with RS of the following 3 algorithms: breast cancer prognostic score, Magee0, and Magee2.**

Design.—**Breast cancer cases with available RSs were reviewed (n = 442). RS categories were stratified by pathologic and biomarker variables. Histopathologic and biomarker data were abstracted from pathology reports, and RS was calculated by each model. Correlation and concordance between models and RS were calculated.**

Results.—**Less than 5% of breast cancers with lobular features, low-grade tumors, carcinomas with high progesterone receptor content, or luminal A tumors had an RS greater than 25. Breast cancer prognostic score, Magee0, and Magee2 demonstrated correlation coefficients with RS of 0.63, 0.61, and 0.62, respectively. Two-step discordances were uncommon. When an RS of 25 was used to separate high-risk from non–high-risk cases, concordance rates of 86% to 88% were achieved.**

Conclusions.—**High RS was observed only in a small percentage of pure or mixed lobular carcinomas, low-grade or luminal A tumors, and tumors with high progesterone receptor expression, suggesting that these cancers may not require Oncotype testing. All 3 surrogate models demonstrated comparable correlation and high concordance with the RS when a cutoff of 25 was used, suggesting their utility in cases where the actual RS is unavailable.**

(*Arch Pathol Lab Med.* 2022;146:1258–1267; doi: 10.5858/arpa.2021-0367-OA)

Breast cancer is the most common cancer in women. Several multigene-based assays have been developed to predict the outcome of breast cancer and guide adjuvant systemic therapy. The Oncotype DX Recurrence Score (RS) was the first commercial assay and was introduced and validated in 2004.[1,2] At present, estrogen receptor (ER)-positive, human epidermal growth factor receptor 2 (HER2)-negative breast carcinoma patients with N0 or N1 disease qualify for Oncotype DX testing. The test assigns an RS between 0 and 100, based on reverse transcriptase dependent quantitative analysis of 21 genes performed on a formalin-fixed, paraffin-embedded tumor section. The risk of distant recurrence and the potential benefit from chemotherapy are proportional to the reported RS. Initially, 18 and 30 were the conventionally used cutoffs to separate low-risk from intermediate-risk categories and intermediate-risk from high-risk categories, respectively.[1,2] More recently, the Trial Assigning Individualized Options for Treatment (TAILORx) validated the use of 11 and 25 as more appropriate cutoffs to separate the 3 risk categories. All women in the low-risk and intermediate-risk groups were found not to benefit from adjuvant chemotherapy.[3,4] There was some benefit of chemotherapy in women 50 years and younger with an RS of 16 to 25.[4] It is well established that the RS influences adjuvant therapy decisions in a significant proportion of early-stage breast cancer patients.[5]

The cost and limited access to Oncotype DX testing by large numbers of breast cancer patients worldwide triggered the development of different surrogate models that aim to predict RS at no extra cost by using existing histopathologic and biomarker features of the tumor. Multiple algorithms have been developed that incorporate various pathologic and biomarker data using the Oncotype DX RS as a gold standard, with favorable results.[6–8] Some of these models use (modified) H-scores for ER and progesterone receptor (PR) levels, while others have used the Allred scoring system. Whether either way of quantifying hormone receptor expression is associated with better correlation of the respective algorithms with the actual RS currently is unclear. Moreover, we are not aware of any study comparing the performance of multiple different surrogate

Accepted for publication September 23, 2021.
Published online February 4, 2022.

From the Department of Pathology and Laboratory Medicine (Mohamed, Kousar, Geradts), Department of Surgery (Wong, Vohra), and the Department of Medical Oncology (Muzaffar), East Carolina University Brody School of Medicine, Greenville, North Carolina.

The authors have no relevant financial interest in the products or companies described in this article.

This study was presented in part at the virtual 2021 United States and Canadian Academy of Pathology Annual Meeting, March 17th, 2021.

Corresponding author: Joseph Geradts, MD, Department of Pathology and Laboratory Medicine, East Carolina University Brody School of Medicine, 600 Moye Blvd, Mailstop 642, Greenville, NC 27834 (email: geradtsj20@ecu.edu).

Table 1. The 3 Predictive Models	
	Calculation Formula
BCPS[7]	RS = 40.0 − 5.3 (ER AS) − 2.7 (PR AS) + 13.0 (1 for HER2 negative, 2 for equivocal, 3 for HER2 positive) + 2.3 (TG) + 2.4 (NG) + 6.5 (MG)
Magee0[6]	RS = 13.424 + 5.420 (NG) + 5.538 (MG) − 0.045 (ER IHC score) − 0.030 (PR IHC score) + 9.486 (0 for HER2 negative, 0.5 for equivocal, 1 for HER2 positive)
Magee2[8]	RS = 18.8042 + 2.34123 NS − 0.03749 (ER IHC score) − 0.03065 (PR IHC score) + (0 for HER2 negative, 1.82921 for equivocal, 11.51378 for HER2 positive) + 0.04267 (tumor size in cm)

Abbreviations: AS, Allred score; BCPS, breast cancer prognostic score; ER, estrogen receptor; HER2, human epidermal growth factor receptor 2; IHC, immunohistochemistry; MG, mitotic grade; NG, nuclear grade; NS, Nottingham score; PR, progesterone receptor; RS, recurrence score; TG, tubular grade.

models in predicting the RS using the updated RS cutoffs based on the TAILORx study.

In the current study, we assembled a large cohort of early stage, ER-positive breast cancer cases to identify individual pathobiologic features that predict a high versus non-high RS and to compare the predictive power of 3 models that are independent of Ki-67 and that incorporate different measures of hormone receptor quantitation.

MATERIALS AND METHODS

Approval of this study was obtained from the East Carolina University Brody School of Medicine institutional review board. A retrospective review of 442 breast resection specimens of ER-positive invasive breast carcinomas received at our institution between 2005 and 2020 with available Oncotype DX RS was performed. The surgical pathology report of each case was reviewed to retrieve Oncotype DX RS, Oncotype DX biomarker subscores, tumor size, histologic type, overall tumor grade, tubular grade, nuclear grade, mitotic grade, combined Nottingham Score, lymph node (LN) status, biomarker profile, including percentage and average intensity of tumor cells staining by ER and PR immunohistochemical (IHC) assays, HER2 IHC staining results, and HER2 fluorescence in situ hybridization reports. Most of the granular HER2 data were not available for an updated classification of HER2 status based on the most recent (2018) College of American Pathologists Clinical Practice Guideline Focused Update. Allred scores and modified H-scores were calculated using ER and PR IHC percentage and intensity of staining.[9,10] Furthermore, the number of the individual block sent for Oncotype DX RS testing and the number of the block used for biomarker testing were noted. Biomarker glass slides of cases for which ER or PR IHC staining intensity and/or percentage was not reported (152 cases) were reviewed and centrally graded by a board-certified anatomic pathologist with expertise in breast pathology (J.G.) (including 111 in-house cases and 41 biopsy specimen cases that were requested from an outside lab). The RS for each case was calculated by the following 3 surrogate models that do not include Ki-67 (since Ki-67 is not routinely performed at our institution): breast cancer prognostic score (BCPS),[7] Magee0,[6] and Magee2[8] (Table 1). The categorical concordance (low versus intermediate versus high) between each model and the Oncotype DX RS was calculated using conventional cutoffs (18, 30) and TAILORx cutoffs (11, 25). In addition, categorical concordance between each model and the Oncotype DX RS was calculated in a dichotomous fashion (RS >25 versus RS ≤25). Cases with 2-step discordance were further investigated and selected glass slides were reviewed, including hematoxylin-eosin–stained sections from paraffin blocks used for Oncotype DX testing and biomarker IHC testing, in addition to ER, PR, and HER2 IHC slides. The correlation coefficient of each model with the Oncotype DX RS was calculated using the Pearson correlation function of Microsoft Excel 2016. Furthermore, cases were categorized into luminal A (low or intermediate-grade tumors with ≥10% ER and PR IHC staining and negative HER2) and luminal B (high-grade tumors, tumors with <10% ER or PR IHC staining, or positive/equivocal HER2). The mean and SD of Oncotype DX RS, as well as the predicted RS for luminal A and luminal B cases, were calculated using Microsoft Excel 2016. Moreover, cases were categorized based on LN status into 3 categories (cases with no LN metastasis, cases with metastasis to 1 LN, and cases with metastasis to 2 LNs).

RESULTS

Of 442 cases, 340 cases (77%) were invasive ductal carcinomas, 47 cases (11%) were invasive lobular carcinomas, 22 cases (5%) had mixed ductal and lobular features, 10 cases (2%) were mucinous carcinomas, and 23 cases had other histologies. One hundred forty-three cases (32%) were low grade, 239 cases (54%) were intermediate grade, and 60 cases (14%) were high grade. Using IHC, all cases were ER positive, 403 cases (91%) were PR positive, and 428 cases (97%) were HER2 negative. Three hundred seventeen cases (72%) were luminal A subtype, while 125 cases (28%) were luminal B subtype (Table 2).

Histopathologic and Biomarker-Based Risk Stratification

Fifty-two of 340 (15%) invasive ductal carcinomas had an RS greater than 25, compared with only 2 of 47 cases (4%) of invasive lobular carcinomas and no mixed carcinomas. Two of 10 mucinous carcinomas had a high RS. Only 7 of 143 (5%) low-grade tumors had an RS greater than 25, compared with 28 of 239 (12%) intermediate grade and 25 of 60 (42%) high-grade tumors. Among tumors with high PR expression, only 13 of 306 (4%) and 10 of 248 (4%) cases had an RS greater than 25 when an Allred score of 7 or higher or a modified H-score of 200 or more was used as a threshold, respectively. Forty-two of 125 cases (34%) of luminal B tumors had an RS greater than 25, compared with only 16 of 317 cases (5%) of luminal A tumors (Table 2). Of note, no statistically significant differences in the RS distribution were observed between LN-negative cases, cases with 1 positive LN, and cases with 2 positive LNs (Figure 1).

IHC-Based Breast Cancer Biomarker Profile Versus Oncotype DX Subscores

A comparison of ER, PR, and HER2 results between IHC and Oncotype DX biomarker subscores demonstrated overall concordance rates of 99.5%, 90%, and 96%, respectively (Table 3). Two ER-positive cases had a negative Oncotype subscore. Thirty-four PR-positive cases had a negative Oncotype subscore, while 4 PR-negative cases were called positive by Oncotype analysis. For HER2, 2 positive and 8 equivocal cases had a negative Oncotype subscore, while 4 negative cases were classified as equivocal by the Oncotype assay (Table 3).

Comparison of Surrogate Models

Luminal B tumors had higher RSs than luminal A tumors (Figure 2, A). Similarly, we found statistically significant differences in the mean and SD of the RS calculated by the 3 models between luminal A and luminal B subtypes, with wider separation between the 2 luminal subtypes by BCPS (Figure 2, B through D). Correlation coefficients of BCPS,

Table 2. Stratification of Recurrence Score (RS) Risk Categories by Pathologic and Biomarker Variables

Conventional Risk Category	Low Risk		Intermediate Risk		High Risk
TAILORx Risk Category	Low Risk	Intermediate Risk		High Risk	
Oncotype DX RS	<11	11–17	18–25	26–30	>30
	n (%)	n (%)	n (%)	n (%)	n (%)
Histologic type (n)					
Invasive ductal carcinoma (340)	94 (28)	121 (36)	73 (21)	20 (6)	32 (9)
Invasive lobular carcinoma (47)	10 (21)	27 (57)	8 (17)	1 (2)	1 (2)
Mixed ductal/lobular carcinoma (22)	3 (14)	15 (68)	4 (18)	0 (0)	0 (0)
Invasive mucinous carcinoma (10)	4 (40)	4 (40)	0 (0)	1 (10)	1 (10)
Molecular subtype (n)					
Luminal A (317)	112 (35)	143 (45)	46 (15)	11 (3)	5 (2)
Luminal B (125)	9 (7)	33 (26)	41 (33)	10 (8)	32 (26)
Combined tumor grade (n)					
Low (143)	45 (31)	59 (41)	32 (22)	4 (3)	3 (2)
Intermediate (239)	69 (29)	101 (42)	41 (17)	13 (5)	15 (6)
High (60)	5 (8)	17 (28)	13 (22)	5 (8)	20 (33)
ER (n)					
High (Allred score 7–8) (408)	120 (29)	164 (40)	74 (18)	21 (5)	29 (7)
High (modified H-score ≥200) (320)	101 (32)	133 (42)	54 (17)	15 (5)	17 (5)
Low (Allred score 3–6) (34)	1 (3)	12 (35)	12 (35)	1 (3)	8 (24)
Low (modified H-score <200) (122)	18 (15)	45 (37)	33 (27)	7 (6)	19 (16)
PR (n)					
High (Allred score 7–8) (306)	106 (35)	146 (48)	41 (13)	9 (3)	4 (1)
High (modified H-score ≥200) (248)	96 (39)	111 (45)	31 (13)	8 (3)	2 (1)
Low (Allred score 3–6) (97)	12 (12)	27 (28)	32 (33)	7 (7)	19 (20)
Low (modified H-score 1–199) (155)	23 (15)	61 (39)	41 (26)	9 (6)	21 (14)
Negative (Allred score 0–2) (39)	1 (3)	5 (13)	14 (36)	5 (13)	14 (36)
Negative (modified H-score <1) (39)	1 (3)	5 (13)	14 (36)	5 (13)	14 (36)
HER2 (n)					
Negative (428)	120 (28)	170 (40)	84 (20)	20 (5)	34 (8)
Positive/equivocal (14)	0 (0)	6 (43)	3 (21)	2 (14)	3 (21)

Abbreviations: ER, estrogen receptor; HER2, human epidermal growth factor receptor 2; PR, progesterone receptor; TAILORx, Trial Assigning Individualized Options for Treatment.

Magee0, and Magee2 with Oncotype DX RS were 0.63, 0.61, and 0.62, respectively. Two-step discordances were uncommon, especially with the TAILORx cutoffs, with the lowest rate of 2-step discordance observed with Magee2 (Figure 3, A through F). Across 3 risk categories (low, intermediate, high), BCPS showed the best categorical concordance with

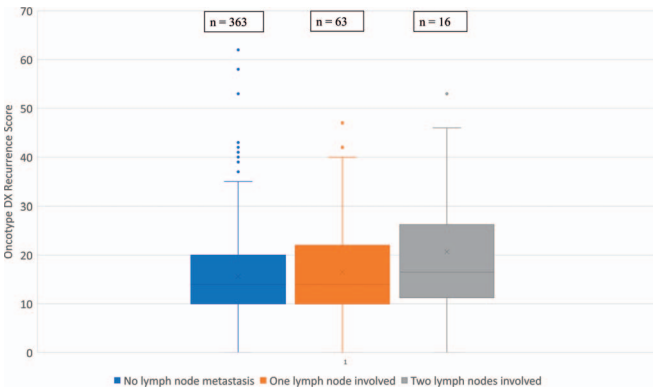

Figure 1. *Box plots of Oncotype DX Recurrence Score of cases with and without lymph node metastases.*

Table 3. Concordance Between Clinical Estrogen Receptor (ER), Progesterone Receptor (PR), and Human Epidermal Growth Factor Receptor 2 (HER2) Status and Oncotype DX Subscore

	Oncotype DX Subscores		
Clinical Lab	Negative	Borderline	Positive
ER			
Negative	0	N/A	0
Positive	2	N/A	379
PR			
Negative	29	N/A	4
Positive	34	N/A	314
HER2			
Negative	365	4	0
Equivocal	8	0	0
Positive	2	2	0

Abbreviation: N/A, not available.

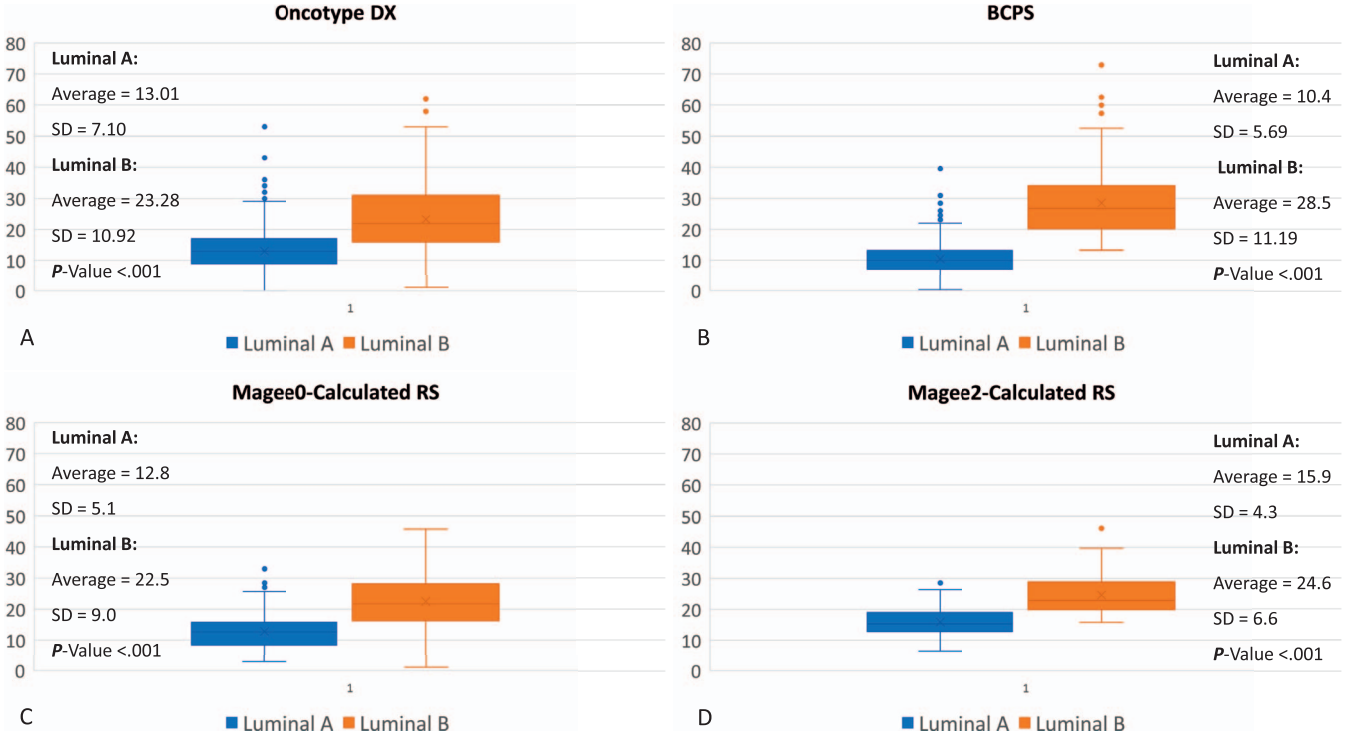

Figure 2. *Box plots showing distribution of Oncotype DX Recurrence Scores (RSs) (A) and calculated model scores (B through D): luminal A versus luminal B tumors. Abbreviation: BCPS, breast cancer prognostic score.*

RS (73%) using conventional cutoffs and 53% concordance using TAILORx cutoffs. Both Magee models yielded similar and slightly better categorical concordance with RS (58%) than BCPS when TAILORx cutoffs were used (Figure 4). The categorical concordance was significantly better for all 3 models when scores were dichotomized as high (>25) versus non-high (≤25), with agreement of 86% to 88% (Figure 5).

Discordant Cases

There were fewer cases with 2-step discordance with TAILORx cutoffs than with conventional cutoffs (Table 4). While only a small number of cases demonstrated 2-step discordance between Oncotype DX RSs and model-calculated scores, none of these cases were 2-step discordant among all the models. Moreover, in many cases the 3 models yielded different predicted RSs. The 2-step discordant cases included both those where different blocks were used for clinical biomarker and Oncotype testing, and those where the same block was used. Upon reviewing the 2-step discordant cases, a few possible explanations emerged. For instance, submitting a paraffin block with low tumor cellularity and dense lymphocytic infiltrate for Oncotype testing resulted in an RS of 27; however, the calculated scores were between 3 and 11 (Figure 6, A through E) (Table 4, case 16). Moreover, the Oncotype ER subscore was low positive, and the Oncotype PR subscore was negative, in contrast to the IHC, which showed a modified H-score of 285 for both ER and PR IHC. Conversely, a high-grade tumor with morphologic intratumoral heterogeneity had a reported RS of 2 when a better differentiated/lower grade area of the tumor was submitted for Oncotype testing, while the models predicted an RS between 25 and 30, partly based on the high overall tumor grade (Figure 7, A through E) (Table 4, case 12).

Finally, we hypothesized that cases where different blocks were used for biomarker and Oncotype testing may show worse correlation and concordance between actual and predicted RSs. However, as shown in Table 5, this was not the case. In fact, in most instances the correlation and concordance looked better for the cases for which different blocks were used.

DISCUSSION

The Oncotype DX Assay has become well established as an important tool in the management of patients with early-stage ER-positive, HER2-negative breast carcinomas. Its prognostic value in patients receiving adjuvant endocrine therapy and its predictive utility for guiding adjuvant chemotherapy have been well validated.[1,2] Other multigene assays, such as MammaPrint and Prosigna (PAM50), provide comparable information.[11,12] However, a significant limitation is the high cost of these assays, and they may not be available in many parts of the world where health care resources are in short supply. To a significant extent, the multigene assays measure the expression of genes that are part of the cancer cells' hormone response and proliferation pathways. Routine pathology reports include data on tumor type and grade as well as ER, PR, and HER2 expression, and a number of models have been developed that amalgamate various pathologic and biomarker parameters in algorithms that correlate well, but not perfectly, with the Oncotype RS as the presumed gold standard. Thus far, only a few reports have compared the performance characteristics of algorithms developed at different institutions.[13,14] The numbers of cases in those studies are smaller than the number of breast carcinomas in our current series (n = 442). An

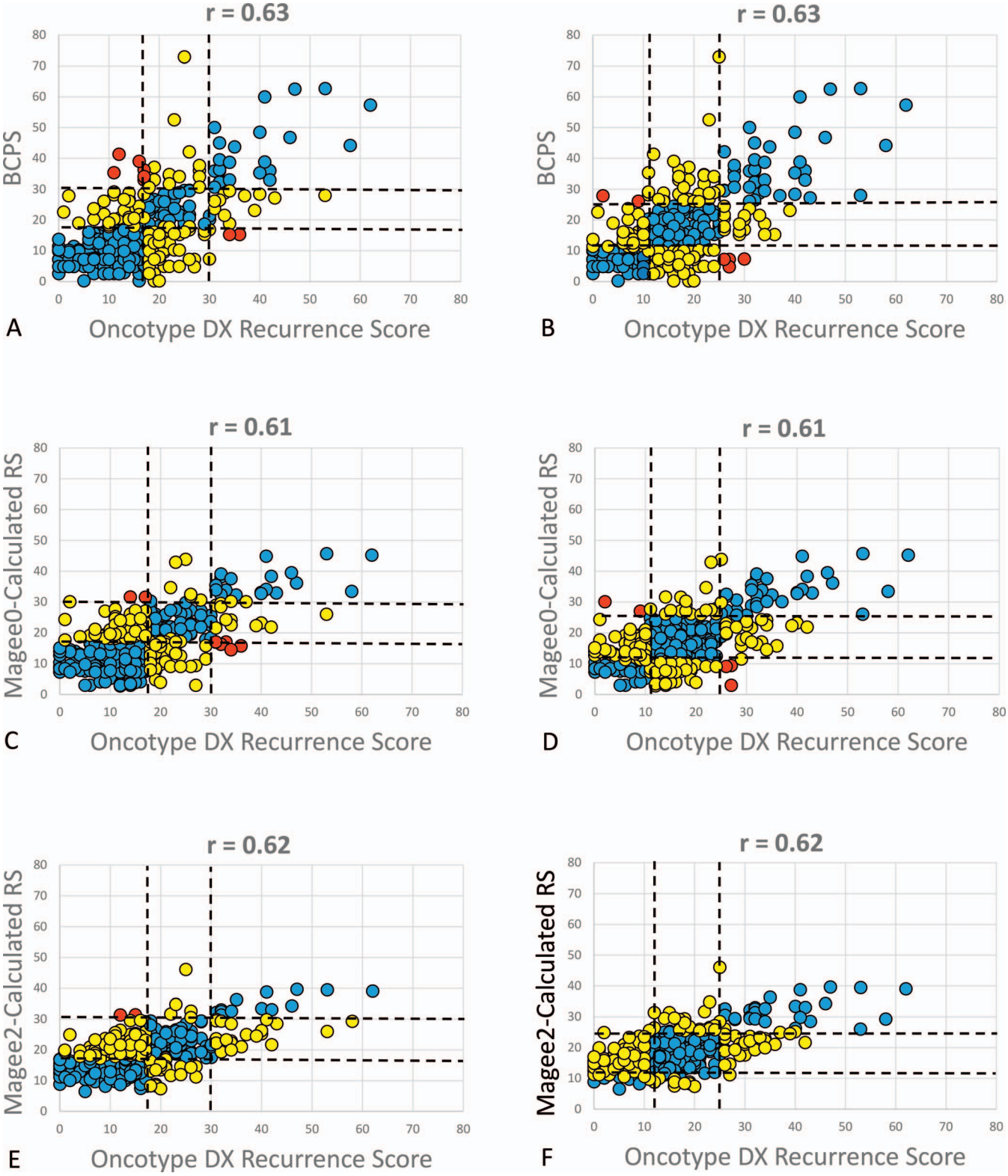

Figure 3. *Scatterplots showing correlation between actual and calculated recurrence scores (RSs) using conventional cutoffs (A, C, and E) and TAILORx cutoffs (B, D, and F). Color coding: blue, categorically concordant cases; yellow, 1-step discordances; red, 2-step discordances. Top panels, breast cancer prognostic score; middle panels, Magee0; bottom panels, Magee2. Abbreviations: BCPS, breast cancer prognostic score; TAILORx, Trial Assigning Individualized Options for Treatment.*

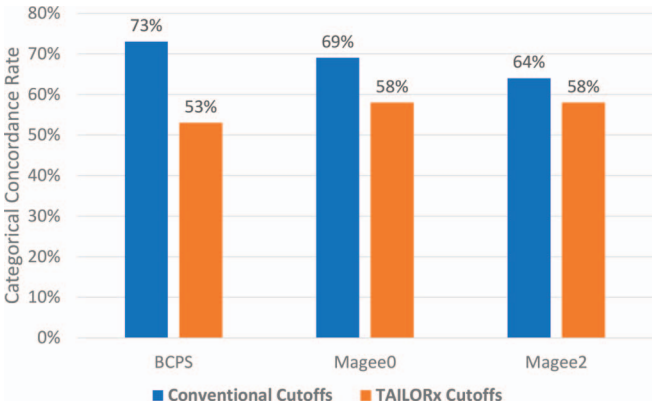

Figure 4. *Categorical concordance of each model with Oncotype DX Recurrence Score using conventional cutoffs and TAILORx cutoffs (3 risk groups: low, intermediate, high). Abbreviations: BCPS, breast cancer prognostic score; TAILORx, Trial Assigning Individualized Options for Treatment.*

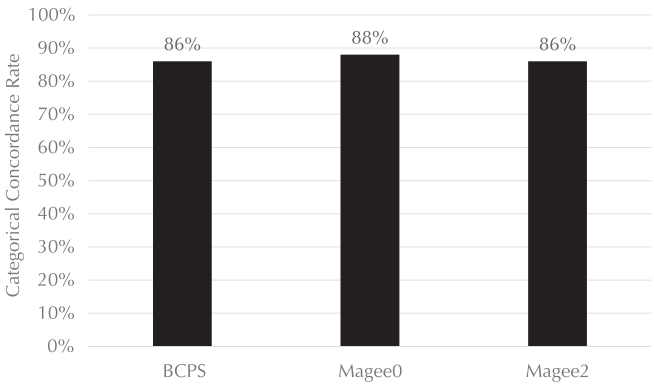

Figure 5. *Categorical concordance of each model with Oncotype DX Recurrence Score—two risk categories (≤25 versus >25). Abbreviation: BCPS, breast cancer prognostic score.*

important goal of our study was to perform comparative validation of some of the earliest models (ie, the Magee equations[6,8] and the BCPS[7]), which use different approaches to quantitating hormone receptor expression.

Tumor histology is relevant to predicting the RS category. In our series of 442 breast carcinomas, only 2 of 47 (4%) lobular carcinomas and none of 22 mixed-ductal/lobular carcinomas had an RS greater than 25. Similar data were reported in previous studies.[7,15,16] Similarly, only 7 of 143 (5%) tumors with a combined grade of 1 and 10 of 248 tumors or 13 of 306 (4%) tumors with high PR content (modified H-score ≥200 or Allred score 7–8, respectively) had an RS greater than 25, confirming other reports.[7,16–18] It was previously shown that ER-positive tumors can be stratified into luminal A and luminal B subtypes using expression levels of ER, PR, HER2, and Ki-67 (or, alternatively, tumor grade).[19] The immunohistochemical subtypes correlate variably well with the molecularly defined subtypes. Concordance is high for basal-like tumors, while it is less than 50% for HER2-enriched tumors.[19] For luminal A and B tumors, concordance rates were reported to be around 70%.[19] Using those criteria, only 16 of 317 (5%) luminal A tumors had an RS greater than 25. Thus, the Oncotype DX assay may not need to be ordered in the classes of tumors mentioned above, because a high RS would be very unlikely. In current practice, HER2-positive breast tumors typically are not submitted for Oncotype testing. However, that was not the case in the first decade after the assay was introduced, and both the BCPS and the Magee equations include HER2 status as a parameter. The definition of HER2-positive and -equivocal status by IHC and fluorescence in situ hybridization has changed several

Table 4. 2-Step Discordant Cases Between Oncotype DX Recurrence Risk Category and Surrogate Models

Case No.	Same Block[a]	Biomarker Block Source	Oncotype Block Source	Oncotype RS	BCPS	Magee0	Magee2
Conventional cutoffs							
1	No	Biopsy	Excision	11	**35**	23	25
2	No	Biopsy	Excision	12	**41**	24	**31**
3	No	Biopsy	Excision	14	27	**32**	29
4	No	Biopsy	Excision	16	**39**	25	30
5	Yes	Biopsy	Biopsy	17	**33**	23	23
6	Yes	Excision	Excision	17	**34**	21	26
7	No	Biopsy	Excision	17	**36**	**32**	29
8	No	Biopsy	Excision	31	26	**17**	22
9	No	Biopsy	Excision	32	27	**16**	19
10	No	Biopsy	Excision	33	29	**17**	22
11	No	Biopsy	Excision	34	**15**	**15**	20
TAILORx Cutoffs							
12	No	Excision	Excision	2	**28**	**30**	25
13	Yes	Excision	Excision	9	**26**	16	22
14	No	Excision	Excision	9	22	**27**	24
15	No	Biopsy	Excision	26	**7**	**9**	14
16	No	Biopsy	Excision	27	**5**	**3**	11
17	No	Biopsy	Excision	30	**7**	18	18

Abbreviations: BCPS, breast cancer prognostic score; RS, recurrence score; TAILORx, Trial Assigning Individualized Options for Treatment.

2-step discordances highlighted in bold.

[a] Same block used for clinical biomarker studies and Oncotype DX assay.

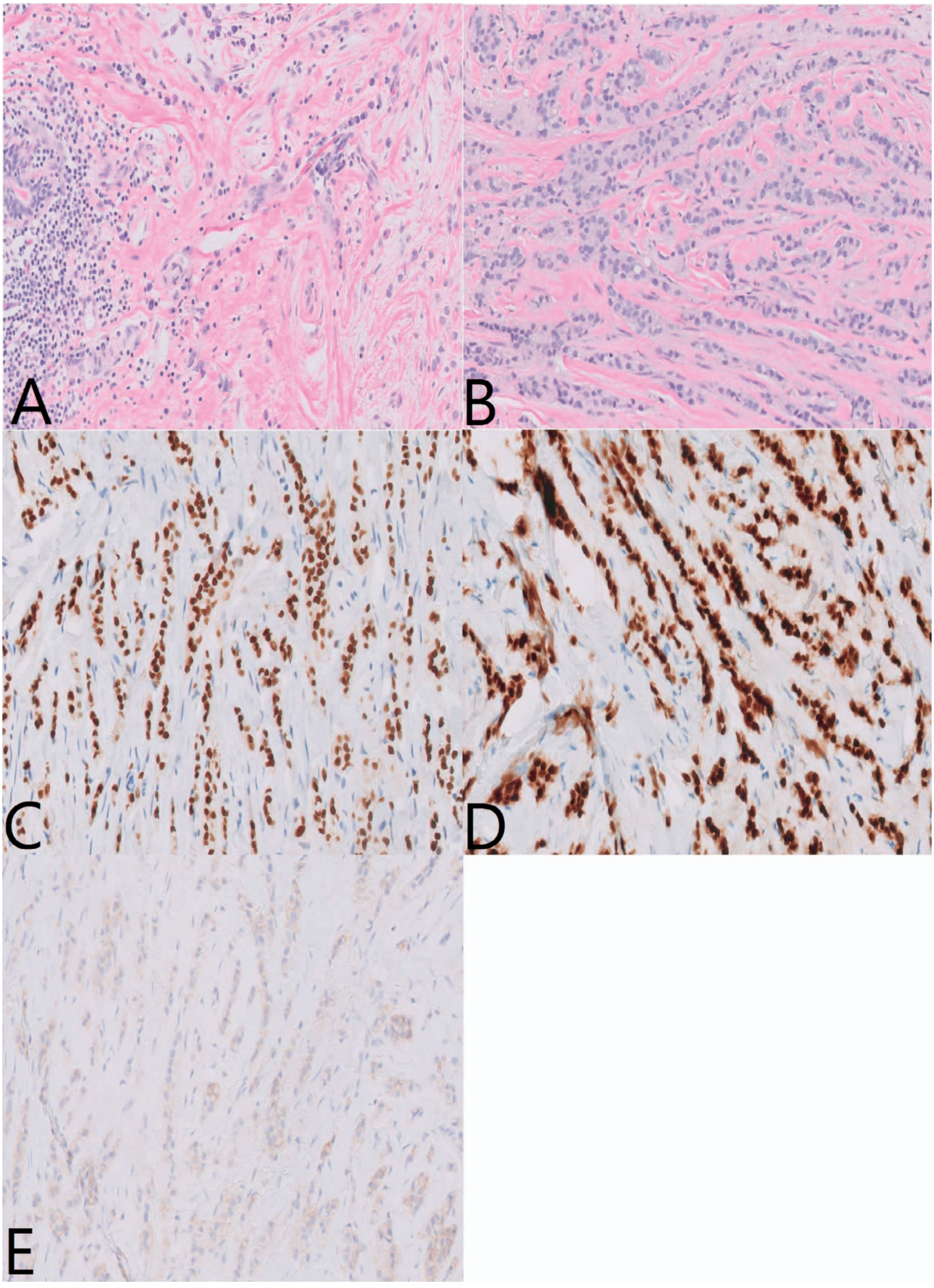

Figure 6. *Discordant case (high recurrence score, low predicted score). Invasive lobular carcinoma with high Oncotype DX Recurrence Score but low model-calculated recurrence scores (case 16 in Table 4). Hematoxylin-eosin staining demonstrates low tumor cellularity and a focally dense lymphoid infiltrate in the block used for Oncotype testing (A). Hematoxylin-eosin staining shows high tumor cellularity and absence of inflammatory cells in the block used for immunohistochemistry (B). The tumor has a modified H-score of 285 for both estrogen receptor (C) and progesterone receptor (D). Tumor cells are negative for human epidermal growth factor receptor 2 by immunohistochemistry (E) (original magnification ×200).*

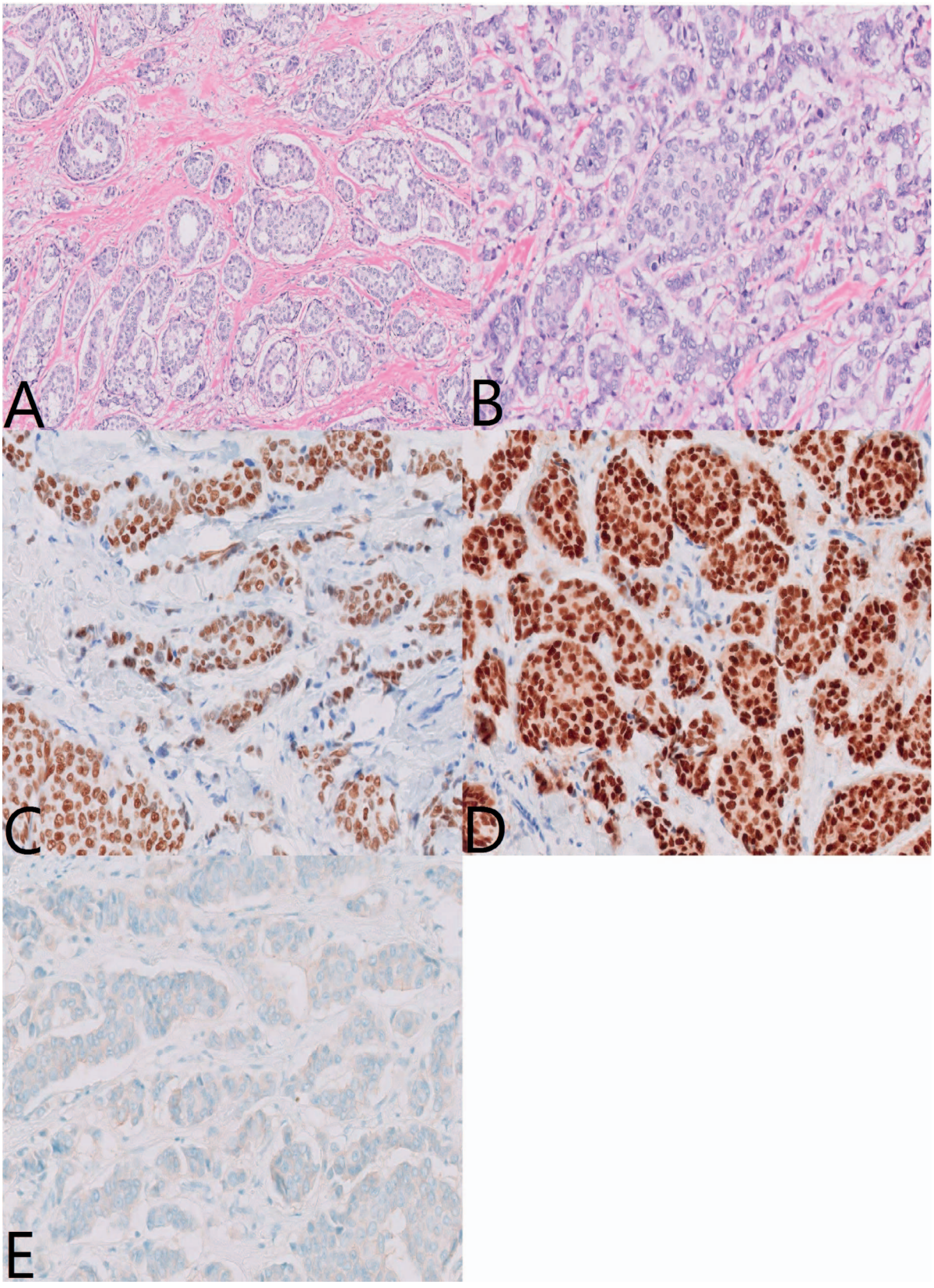

Figure 7. *Discordant case (low recurrence score, high predicted score). High-grade invasive ductal carcinoma with low Oncotype DX Recurrence Score but high model-calculated recurrence scores (case 12 in Table 4). Hematoxylin-eosin staining demonstrates an area of better differentiation and lower nuclear grade in the block used for Oncotype testing (A). Hematoxylin-eosin staining shows solid nests of tumor cells with high nuclear grade and increased mitotic activity in the block used for biomarker immunohistochemistry (B). The tumor has an estrogen receptor modified H-score of 160 (C) and a progesterone receptor H-score of 300 (D). The tumor cells are negative for human epidermal growth factor receptor 2 by immunohistochemistry (E) (original magnification ×200).*

Table 5. Correlation and Concordance Rates Between Surrogate Models and Oncotype Recurrence Score (RS) When the Same or Different Blocks Were Used for Oncotype and Clinical Biomarker Testing

	BCPS	Magee0	Magee2
Correlation coefficient (*r*)			
Combined	0.63	0.61	0.62
Same block[a]	0.52	0.56	0.5
Different block[b]	0.66	0.62	0.64
Categorical concordance: 3 categories (conventional cutoffs)			
Combined	73%	69%	64%
Same block	74%	63%	61%
Different block	73%	72%	67%
Categorical concordance: 3 categories (TAILORx cutoffs)			
Combined	53%	58%	58%
Same block	47%	54%	52%
Different block	55%	58%	60%
Categorical concordance: 2 categories (≤25 versus >25)			
Combined	86%	88%	86%
Same block	83%	91%	86%
Different block	87%	88%	88%

Abbreviations: BCPS, breast cancer prognostic score; TAILORx, Trial Assigning Individualized Options for Treatment.

[a] Cases with same block for immunohistochemical (IHC) biomarker and Oncotype testing: n = 40.

[b] Cases with different blocks for IHC biomarker and Oncotype testing: n = 402.

times between 2004 and 2020, and we abstracted the HER2 data based on guidelines that were in place at the respective times.

It was previously reported that the Oncotype DX subscores for ER, PR, and HER2 expression generally correlated well with the clinical biomarker data.[7] More recently, 2 larger studies confirmed that concordance for ER exceeded 98%, while it was somewhat reduced for PR (86.6% and 90%, respectively).[20,21] In our study, only 2 of 391 (0.5%) ER-positive breast carcinomas were called negative in the Oncotype DX subscore analysis. Thirty-eight of 381 (10%) of cases were discordant for PR, and 34 of those were positive in the clinical assay and negative by Oncotype analysis, possibly as a result of intratumoral heterogeneity, which is significantly more common for PR than for ER.[22] In our study, only 2 tumors were positive, and 8 were equivocal for HER2 in the clinical biomarker assays, and all of them had a negative HER2 subscore by Oncotype analysis. Similar observations were previously made by other investigators.[20,21] At present, there is insufficient evidence to suggest that the Oncotype DX subscores for ER, PR, and HER2 could be used to guide endocrine or anti-HER2 therapy.

There are a number of models seeking to predict the Oncotype RS. Many of them include Ki-67 as an important marker for tumor proliferation. However, marked variability exists in the sensitivity of Ki-67 stains performed in different laboratories and in the interpretation of those stains.[23] In our opinion, this is a significant limitation of algorithms that incorporate Ki-67. Thus, we decided to perform comparative validation studies for several models that do not rely on Ki-67, specifically the BCPS,[7] the original Magee equation,[6] and Magee equation 2.[8] Unlike the former 2 algorithms, the latter also includes tumor size as a variable, which can be problematic in needle core biopsy specimens. The 2 Magee equations use the H-score for ER and PR, while subsequent modified versions of the Magee equations employ the modified H-score. The (modified) H-score has a wider dynamic range (0–300) and thus may arguably be less reproducible than the Allred score that is used to compute the BCPS. In our series of 442 breast carcinomas, all 3 models showed very similar correlation with the Oncotype RS ($r = 0.61$–0.63). This observation is remarkably similar to previous reports.[10,13,14,16,24] With regard to concordance across 3 categories (low, intermediate, high), agreement ranged from 64% to 73% for the conventional cutoffs and from 53% to 58% for the TAILORx cutoffs. This may partly be explained by the fact that the BCPS and Magee equations were developed for optimal concordance of the respective model with the Oncotype RS using the thresholds that were generally used at the time. Similar concordance rates were previously published.[7,16,24,25] Of note, Magee equation 2 did not perform appreciably better than the original Magee equation that had been published 5 years earlier and that excludes tumor size as a variable (Figures 2 through 5). After the results of the TAILORx trial were published,[3,4] most clinicians adopted an RS threshold of more than 25 to determine potential chemotherapy benefit, at least in postmenopausal women. When this newer threshold was used to dichotomize our cases (high versus non-high), concordance rates improved to almost 90%. This suggests the possibility that the 3 surrogate models can be considered in clinical decision making if actual Oncotype RS data are not available. Of note, our study demonstrates the equivalence of algorithms that use either Allred or (modified) H-scores for ER/PR quantitation. To our knowledge this is one of the few studies that has used the updated RS cutoffs based on the TAILORx trial to assess the concordance with multiple surrogate models.

We sought to understand the possible reasons for discordance between actual and predicted Oncotype RSs. Undoubtedly, methodologic issues are important. The Oncotype DX assay measures mRNA expression of 16 cancer and 5 reference genes in a homogenate that includes a multitude of different cell types in variable and unknown proportions, in addition to the neoplastic cells (that by themselves may be a mixture of in situ and invasive cancer cells). In contrast, clinical ER, PR, and HER2 assays specifically quantify individual protein expression in invasive tumor cells; in the case of HER2, a fluorescence in situ hybridization assay for gene amplification may be used as an alternative. Consequently, these assays are not impacted by variable tumor cellularity. Tumor grading, which has suboptimal interobserver reproducibility, is based on examining all blocks containing invasive carcinoma. For the BCPS, it was previously shown that the concordance rate did not improve with central grading.[7] In the current study, 2-step discordances were uncommon, especially with the TAILORx cutoffs (Table 5), consistent with other reports.[7,10,25] Surprisingly, correlation coefficients and concordance rates for all 3 models were not worse when different blocks were used for clinical biomarker and Oncotype DX assays; in fact, the numbers seemed more favorable (Table 5). This may be a reassuring observation and suggests that there may not be an advantage to using the same tumor block for both targeted biomarker and Oncotype DX assays. In some instances, an unexpectedly high RS might be due to reduced tumor cellularity (ie, fewer cells with an activated ER signaling

pathway) or admixture of an inflammatory infiltrate (Figure 6). It was previously reported that inflammatory cells may elevate the RS by increasing the proliferation score.[26,27] Moreover, a needle core biopsy specimen can induce inflammatory changes that increase proliferation in the subsequent excisional specimen.[28] Discordances in actual and computed RSs may also be due to intratumoral heterogeneity.[22] Both biomarker expression and pathologic features including proliferative activity, nuclear grade, and degree of differentiation may vary significantly within a given tumor. As illustrated in Figure 7, an RS may be unexpectedly low if a lower grade area of a tumor is selected for Oncotype testing. Thus, it may be advisable to select a block that is representative of the overall tumor grade, ideally including the mitotically most active area.

CONCLUSIONS

Less than 5% of breast carcinomas with pure or mixed lobular differentiation, low combined grade, high PR content, or luminal A subtype have an Oncotype RS greater than 25, and thus these tumors may not require genomic risk assessment by the Oncotype DX assay. The 3 surrogate models (BCPS, Magee0, Magee2) had comparable correlation coefficients and concordance rates with the reported RS, suggesting that, for ER and PR, Allred and modified H-scores are similarly informative. All 3 models predicted an RS greater than 25 with high (≥86%) accuracy, and thus they may be useful in managing breast cancer patients for whom actual Oncotype DX data are not available. Possible reasons for discordance in predicted versus actual RS include variable tumor cellularity, inflammatory infiltrates, intratumoral heterogeneity, and methodologic differences.

We gratefully acknowledge the contribution of William Balance, MD, who generously provided slides and data for a subset of cases in our study.

References

1. Paik S, Shak S, Tang G, et al. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N Engl J Med*. 2004;351(27):2817–2826. doi:10.1056/NEJMoa041588

2. Paik S, Tang G, Shak S, et al. Gene expression and benefit of chemotherapy in women with node-negative, estrogen receptor-positive breast cancer. *J Clin Oncol*. 2006;24(23):3726–3734. doi:10.1200/JCO.2005.04.7985

3. Sparano JA, Gray RJ, Makower DF, et al. Prospective validation of a 21-gene expression assay in breast cancer. *N Engl J Med*. 2015;373(21):2005–2014. doi:10.1056/NEJMoa1510764

4. Sparano JA, Gray RJ, Makower DF, et al. Adjuvant chemotherapy guided by a 21-gene expression assay in breast cancer. *N Engl J Med*. 2018;379(2):111–121. doi:10.1056/NEJMoa1804710

5. Lo SS, Mumby PB, Norton J, et al. Prospective multicenter study of the impact of the 21-gene recurrence score assay on medical oncologist and patient adjuvant breast cancer treatment selection. *J Clin Oncol*. 2010;28(10):1671–1676. doi:10.1200/JCO.2008.20.2119

6. Flanagan MB, Dabbs DJ, Brufsky AM, Beriwal S, Bhargava R. Histopathologic variables predict Oncotype DX recurrence score. *Mod Pathol*. 2008;21(10):1255–1261. doi:10.1038/modpathol.2008.54

7. 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(9):969–977. doi:10.3109/07357907.2010.512600

8. Klein ME, Dabbs DJ, Shuai Y, et al. Prediction of the Oncotype DX recurrence score: use of pathology-generated equations derived by linear regression analysis. *Mod Pathol*. 2013;26(5):658–664. doi:10.1038/modpathol.2013.36

9. Allred DC. Issues and updates: evaluating estrogen receptor-alpha, progesterone receptor, and HER2 in breast cancer. *Mod Pathol*. 2010;23 Suppl 2:S52–S59. doi:10.1038/modpathol.2010.55

10. Turner BM, Skinner KA, Tang P, et al. Use of modified Magee equations and histologic criteria to predict the Oncotype DX recurrence score. *Mod Pathol*. 2015;28(7):921–931. doi:10.1038/modpathol.2015.50

11. Nielsen TO, Parker JS, Leung S, et al. A comparison of PAM50 intrinsic subtyping with immunohistochemistry and clinical prognostic factors in tamoxifen-treated estrogen receptor-positive breast cancer. *Clin Cancer Res*. 2010;16(21):5222–5232. doi:10.1158/1078-0432.CCR-10-1282

12. Cardoso F, van't Veer LJ, Bogaerts J, et al. 70-gene signature as an aid to treatment decisions in early-stage breast cancer. *N Engl J Med*. 2016;375(8):717–729. doi:10.1056/NEJMoa1602253

13. Harowicz MR, Robinson TJ, Dinan MA, et al. Algorithms for prediction of the Oncotype DX recurrence score using clinicopathologic data: a review and comparison using an independent dataset. *Breast Cancer Res Treat*. 2017;162(1):1–10. doi:10.1007/s10549-016-4093-4

14. Robertson SJ, Pond GR, Hilton J, et al. Selecting patients for Oncotype DX testing using standard clinicopathologic information. *Clin Breast Cancer*. 2020;20(1):61–67. doi:10.1016/j.clbc.2019.07.006

15. Tsai ML, Lillemoe TJ, Finkelstein MJ, et al. Utility of Oncotype DX risk assessment in patients with invasive lobular carcinoma. *Clin Breast Cancer*. 2016;16(1):45–50. doi:10.1016/j.clbc.2015.08.001

16. Glasgow A, Sechrist H, Bomeisl P, Gilmore H, Harbhajanka A. Correlation between modified Magee equation-2 and Oncotype-Dx recurrence scores using both traditional and TAILORx cutoffs and the clinical application of the Magee Decision Algorithm: a single institutional review. *Breast Cancer*. 2021;28(2):321–328. doi:10.1007/s12282-020-01163-3

17. Bomeisl PE, Thompson CL, Harris LN, Gilmore HL. Comparison of Oncotype DX recurrence score by histologic types of breast carcinoma. *Arch Pathol Lab Med*. 2015;139(12):1546–1549. doi:10.5858/arpa.2014-0557-OA

18. Hou Y, Zynger DL, Li X, Li Z. Comparison of Oncotype DX with modified Magee equation recurrence scores in low-grade invasive carcinoma of breast. *Am J Clin Pathol*. 2017;148(2):167–172. doi:10.1093/ajcp/aqx059

19. Allott EH, Geradts J, Cohen SM, et al. Frequency of breast cancer subtypes among African American women in the AMBER consortium. *Breast Cancer Res*. 2018;20(1):12. doi:10.1186/s13058-018-0939-5

20. Neely C, You S, Mendoza PM, Aneja R, Sahin AA, Li X. Comparing breast biomarker status between routine immunohistochemistry and FISH studies and Oncotype DX testing, a study of 610 cases. *Breast J*. 2018;24(6):889–893. doi:10.1111/tbj.13110

21. Sechrist H, Glasgow A, Bomeisl P, Gilmore H, Harbhajanka A. Concordance of breast cancer biomarker status between routine immunohistochemistry/in situ hybridization and Oncotype DX qRT-PCR with investigation of discordance, a study of 591 cases. *Hum Pathol*. 2020;104:54–65. doi:10.1016/j.humpath.2020.07.022

22. Allott EH, Geradts J, Sun X, et al. Intratumoral heterogeneity as a source of discordance in breast cancer biomarker classification. *Breast Cancer Res*. 2016;18(1):68. doi:10.1186/s13058-016-0725-1

23. Polley MY, Leung SC, McShane LM, et al. An international Ki67 reproducibility study. *J Natl Cancer Inst*. 2013;105(24):1897–1906. doi:10.1093/jnci/djt306

24. Hou Y, Tozbikian G, Zynger DL, Li Z. Using the Modified Magee Equation to identify patients unlikely to benefit from the 21-Gene Recurrence Score Assay (Oncotype DX Assay). *Am J Clin Pathol*. 2017;147(6):541–548. doi:10.1093/ajcp/aqx008

25. Walts AE, Mirocha JM, Bose S. Comparison of Magee and Oncotype DX recurrence scores in estrogen receptor positive breast cancers. *Breast J*. 2018;24(6):951–956. doi:10.1111/tbj.13108

26. 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 Cancer Res Treat*. 2012:413–424. Doi:10.1007/s10549-011-1416-3

27. Acs G, Esposito NN, Kiluk J, Loftus L, Laronga C. A mitotically active, cellular tumor stroma and/or inflammatory cells associated with tumor cells may contribute to intermediate or high Oncotype DX recurrence scores in low-grade invasive breast carcinomas. *Mod Pathol*. 2012:556–566. doi:10.1038/modpathol.2011.194

28. Szalayova G, Ogrodnik A, Spencer B, et al. Human breast cancer biopsies induce eosinophil recruitment and enhance adjacent cancer cell proliferation. *Breast Cancer Res Treat*. 2016;157(3):461–474. doi:10.1007/s10549-016-3839-3