VOLUME 27 · NUMBER 17 · JUNE 10 2009

JOURNAL OF CLINICAL ONCOLOGY

ORIGINAL REPORT

Comparison of Models to Predict Nonsentinel Lymph Node Status in Breast Cancer Patients With Metastatic Sentinel Lymph Nodes: A Prospective Multicenter Study

Charles Coutant, Camille Olivier, Eric Lambaudie, Eric Fondrinier, Fréderic Marchal, François Guillemin, Nathalie Seince, Véronique Thomas, Jean Levêque, Emmanuel Barranger, Emile Darai, Serge Uzan, Gilles Houvenaeghel, and Roman Rouzier

From the Department of Obstetrics and Gynecology, Hôpital Tenon, Assistance Publique Hôpitaux de Paris and University Pierre et Marie Curie, Paris 6 and UPRES EA 4053; Department of Obstetrics and Gynecology, Hôpital Lariboisière, Assistance Publique Hôpitaux de Paris, Paris; Oncology Surgical Department, Institut Paoli-Calmettes, Marseille; Oncology Surgical Department, Institut Paul Papin, Angers; Oncology Surgical Department, Institut Alexis Vautrin, Nancy; Department of Obstetrics and Gynecology, Hôpital Jean Verdier, Assistance Publique Hôpitaux de Paris, Bondy; Oncology Surgical Department, Institut Eugène Marquis; and Department of Obstetrics and Gynecology, Centre Hospitalier Universitaire Rennes, Rennes France.

Submitted August 27, 2008; accepted January 7, 2009; published online ahead of print at www.jco.org on April 6, 2009.

Authors' disclosures of potential conflicts of interest and author contributions are found at the end of this article.

Corresponding author: Charles Coutant, MD, Department of Obstetrics and Gynecology, Hôpital Tenon, 4 rue de la Chine, 75020 Paris, and University Pierre et Marie Curie, Paris 6 and UPRES EA 4053, Paris, France; e-mail: charles.coutant@tnn.aphp.fr.

The Appendix is included in the full-text version of this article, available online at www.jco.org. It is not included in the PDF version (via Adobe® Reader®).

0732-183X/09/2717-2800/$20.00

DOI: 10.1200/JCO.2008.19.7418

ABSTRACT

Purpose
Several models have been developed to predict nonsentinel lymph node (non-SN) status in patients with breast cancer with sentinel lymph node (SN) metastasis. The purpose of our investigation was to compare available models in a prospective, multicenter study.

Patients and Methods
In a cohort of 561 positive-SN patients who underwent axillary lymph node dissection, we evaluated the areas under the receiver operating characteristic curves (AUCs), calibration, rates of false negatives (FN), and number of patients in the group at low risk for non-SN calculated from nine models. We also evaluated these parameters in the subgroup of patients with micrometastasis or isolated tumor cells (ITC) in the SN.

Results
At least one non-SN was metastatic in 147 patients (26.2%). Only two of nine models had an AUC greater than 0.75. Three models were well calibrated. Two models yielded an FN rate less than 5%. Three models were able to assign more than a third of patients in the low-risk group. Overall, the Memorial Sloan-Kettering Cancer Center nomogram and Tenon score outperform other methods for all patients, including the subgroup of patients with only SN micrometastases or ITC.

Conclusion
Our study suggests that all models do not perform equally, especially for the subgroup of patients with only micrometastasis or ITC in the SN. We point out available evaluation methods to assess their performance and provide guidance for clinical practice.

J Clin Oncol 27:2800-2808. © 2009 by American Society of Clinical Oncology

INTRODUCTION

Sentinel lymph node (SN) biopsy can accurately stage the axilla in early breast cancer, and it causes less morbidity than axillary lymph node dissection (ALND).[1-3] It remains to be determined whether ALND is always required for women with positive SNs on final histology, given that 40% to 70% of these patients have no metastatic nonsentinel lymph nodes (non-SNs).[4-9] The likelihood of non-SN metastasis depends on several factors, such as histologic primary tumor size, the size of SN metastasis, the number of positive SNs, the ratio of positive SNs to all removed SNs, and the extracapsular extension status of the positive SNs.[5,6,9-13] However, none of these characteristics by themselves can identify a subset of patients for whom ALND is unnecessary.

Several mathematical models have been developed to predict non-SN status in patients with breast cancer with SN metastasis. These include four nomograms (ie, the Memorial Sloan-Kettering Cancer Center [MSKCC] nomogram, the nomogram developed by Degnim et al [Mayo nomogram], the nomogram developed by Pal et al [Cambridge nomogram], and the nomogram developed by Kohrt et al [Stanford nomogram]), three scoring systems (ie, the Tenon score, the score from The University of Texas M. D. Anderson Cancer Center [MDA score], and the score developed by Saidi et al [Saidi score]), and two recursive partitioning (RP) tools developed by Kohrt et al.[6,10-12,14-16] However, before being incorporated into routine clinical practice, such models must be validated in independent patient populations.

The aim of this article was to evaluate and compare the several tools in an independent, multicenter cohort of patients with breast cancer and positive SNs. Moreover, results were compared

Table 1. Clinical and Pathologic Data for the 561 Patients With Breast Cancer Having Positive SNs and the Subgroup of Patients With Micrometastases or ITC

Clinical and Pathologic Data	Entire Cohort (N = 561)		Subgroup of Micrometastases or ITC (n = 246)	
	No.	%	No.	%
Age, years				
Mean	55.9		55.9	
Range	25-99		25-99	
Postmenopausal	346	61.7	146	59.3
Palpable mass	421	75	181	73.6
Breast surgery				
Lumpectomy	504	89.8	228	92.7
Mastectomy	57	10.2	18	7.3
Invasive tumor size at final histology, cm				
Mean	1.88		1.67	
Range	0.1-14		0.1-6	
≤ 1	118	21	59	24
1.1-2	278	49.6	140	56.9
2.1-3	97	17.3	28	11.4
3.1-5	60	10.7	18	7.3
> 5	8	1.4	1	0.4
Histology				
Invasive ductal carcinoma	468	83.4	207	84.1
Invasive lobular carcinoma	60	10.7	22	8.9
Other	33	5.9	17	6.9
Tumor grade				
Well differentiated, grade 1	172	30.7	82	33.3
Moderately differentiated, grade 2	260	46.3	114	46.3
Poorly differentiated, grade 3	129	23	50	20.3
Lymphovascular space involvement				
No	368	65.6	178	72.3
Yes	193	34.4	68	27.6
Estrogen/progesterone receptor status				
Positive	507	90.4	219	89
Negative	54	9.6	27	11
Her-2/*neu* status				
Negative	161/193	83.4	79/91	86.8
Positive	32/193	16.6	12/91	13.2
Not determined	368	66	155	63
SNs				
No. of SNs removed	1375		629	
Mean No. of SNs per patient	2.45		2.56	
Range	1-9		1-9	
No. of positive SNs	733		295	
No. of patients with:				
Macrometastases	315	56.1	—	
Micrometastases on H&E	159	28.3	159	28.3
Micrometastases on IHC	50	8.9	50	8.9
ITC	37	6.6	37	6.6
Histologic detection of SN metastasis				
Routine H&E	110	19.6	7	2.8
Serial H&E	268	47.8	154	62.6
IHC only	85	15.2	85	34.5
Not determined	98	17.5	0	
Ratio of metastatic SNs to total SNs				
Mean ratio	0.64		0.5	
Range	0.11-1		0.11-1	
1	201	35.8	69	28
0.5-1	219	39	98	39.8
< 0.5	141	25.1	79	32.1
Extracapsular extension				
Yes	61	10.9	6	2.4
No	500	89.1	240	97.6
Non-SNs				
No. of non-SNs removed	6,034		2,494	
Mean No. of non-SNs per patient	10.8		11.13	
Range	2-37		2-37	
No. of patients with positive non-SNs	147	26.2	20	8.1

Abbreviations: SN, sentinel lymph node; ITC, isolated tumor cells; H&E, hematoxylin and eosin; IHC, immunohistochemistry; non-SN, nonsentinel lymph node.

with an optimal logistic regression (OLR) model that was developed from the patient cohort. We also tested the accuracy of each model in the subgroup of patients with micrometastases or isolated tumor cells (ITC) in the SN.

PATIENTS AND METHODS

Study Population

From January 2004 to December 2007, data on 561 patients with breast cancer and positive SNs were prospectively recorded into a single database; patient data came from six institutions: Tenon Hospital (Paris, France; 47 patients), Paoli-Calmettes Cancer Center (Marseille, France; 223 patients), Paul Papin Cancer Center (Angers, France; 113 patients), Alexis Vautrin Cancer Center (Nancy, France; 51 patients), Jean Verdier hospital (Bondy, France; 15 patients), and Eugène Marquis Cancer Center (Rennes, France; 112 patients). No patients had neoadjuvant treatment before SN biopsy or ALND. All the patients signed an informed consent form. Table 1 describes the characteristics of the population, as well as tumor and histologic findings on 561 patients and on a subgroup of 246 patients with only SN micrometastases or ITC. None of the patients included in the current series had been previously included in a series used to develop a model.

SN biopsy was performed as previously described.[10,17] All patients included in this study had completion ALND. ALND was performed during the same procedure when the SN was positive by imprint cytology or frozen section, or when the primary tumor was more than 2 cm intraoperatively. A second operation was performed when either hematoxylin and eosin staining or immunohistochemistry revealed tumor cells in the SN postoperatively, including ITC. Pathologic SN examination methods were as reported previously.[10,17]

Description of Several Mathematical Models

Nine models developed to predict non-SN metastasis in positive-SN patients using postoperative information were identified in the medical literature using PubMed.[6,10-12,14-16] Table 2 describes each model.

Nomograms predict individual patient risk on the basis of a multivariate logistic regression (LRM) analysis. Probabilities of non-SN involvement were obtained from a dedicated Web site (MSKCC, Stanford nomograms), from a graphical interface (Mayo nomogram), or from a formula (Cambridge nomogram).[11,12,14,15]

Scoring systems are based on attribution of points for informative variables. Results are compared with a threshold that allows the separation of patients into low-risk or high-risk groups. The thresholds for the Tenon score (range, 0 to 7), MDA score (range, −2 to 4), and Saidi score (range, 0 to 5) were 3.5, 0, and 2, respectively.[6,10,16]

Table 2. Description of Models to Predict Nonsentinel Node Involvement in Breast Cancer Patients With Sentinel Node Metastasis

Characteristic	MSKCC Nomogram[12]	Mayo Nomogram[11]	Cambridge Nomogram[15]	Stanford Nomogram[14]	MDA Score[6]	Tenon Score[10]	Saidi Score[16]	RP-ROC[14]	CART[14]
Type	Nomogram	Nomogram	Nomogram	Nomogram	Score (−2 to 4)	Score (0 to 7)	Score (0 to 4)	RP	RP
Threshold	≤ 10%	≤ 10%	≤ 10%	≤ 10%	≤ 0	≤ 3.5	≤ 2	≤ 10%	≤ 10%
Age of patient according to estrogen receptor status		X							
Histologic tumor size		X		X*	0 pt when ≤ 10 mm, 1 pt when > 10 mm	0 pt when ≤ 10 mm, 1.5 pt when 0 to 20 mm, 3 pts when > 20 mm	2 pts when > 10 mm, 1 pt if not	X	X
Palpable mass or not							1 pt, 0 if not		
Tumor type	X								
Histologic grade			X						
Lymphovascular invasion	X			X†	1 pt, 0 if not		1 pt, 0 if not	X	X
Multifocality	X								
Estrogen receptor status	X								
No. of SNs removed					−2 pt when ≥ 3				
No. of negative SNs	X	X							
No. of positive SNs	X	X							
Extracapsular extension		X‡					1 pt, 0 if not		
The size of SN metastasis		X	X§	X†, X*	2 pts when macrometastasis, 0 if not	2 pts when macrometastasis, 0 if not		X	X
Method of detection of SN metastases	X								
Ratio of positive SNs to all removed SNs			X			0 pt when ≤ 0.5, 1 pt when 0.5-1, 2 pts when = 1			

Abbreviations: MSKCC, Memorial Sloan-Kettering Cancer Center; MDA, The University of Texas M. D. Anderson Cancer Center; RP-ROC, recursive partitioning with receiver operating characteristic; CART, boosted Classification and Regression Trees; RP, recursive partitioning; pt, point; SN, sentinel lymph node.
*Composite variable: tumor size × (size of SN metastasis)^2.
†Composite variable: lymphovascular invasion × size of SN metastasis.
‡No. of positive SNs according to the presence or absence of extracapsular extension.
§Overall size of SN metastases.

Two RP models were proposed by Kohrt et al: recursive partitioning with receiver operating characteristic (RP-ROC) and boosted Classification and Regression Trees (CART).[14] RP is nonparametric in nature; it imposes no a priori restrictions on the distributional forms of the predictor variables. The RP algorithm is simple and intuitive. At each step, the RP program determines threshold values for each variable that provide the best separation of patients into homogeneous groups. Subgroups were separated using either the RP-ROC method as described by Kraemer (software available from Sierra-Pacific Mental Illness Research, Education, and Clinical Centers [Palo Alto, CA]), or the CART method.[18,19] In the latter case, splits were chosen using the Gini criterion.[18]

Data and Statistical Analysis

Clinical and pathologic data for each patient were recorded prospectively. The performance of each model was quantified with respect to discrimination, calibration, false negative (FN) rate, and clinical utility.

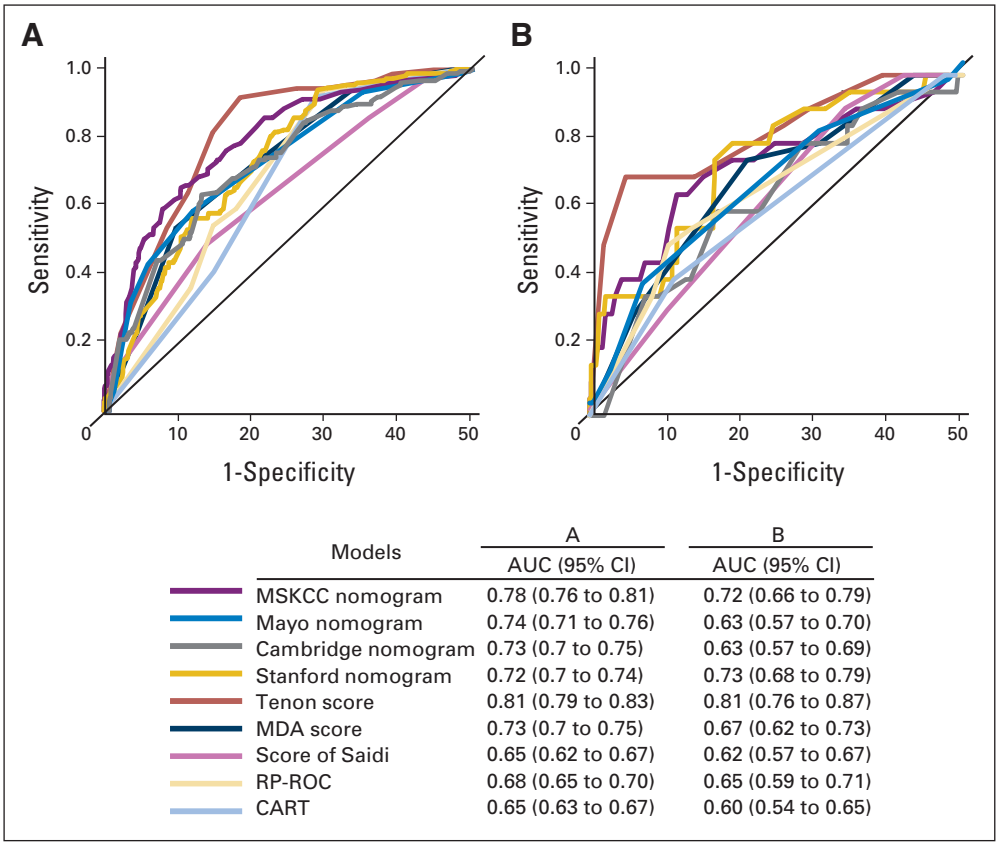
Discrimination. Discrimination (ie, whether the relative ranking of individual predictions is in the correct order) was quantified with the area under the receiver operating characteristic (ROC) curve (AUC).[20] The AUC is a summary measure of the ROC that reflects the ability of a test to discriminate between a diseased and a nondiseased subject across all possible levels of positivity. A 95% CI was calculated for each AUC. AUC ranges from 0 to 1, with 1 indicating perfect concordance, 0.5 indicating no better concordance than chance, and 0 indicating perfect discordance.

Calibration. Calibration (ie, agreement between observed outcome frequencies and predicted probabilities) was studied from graphical representations of the relationship between the observed outcome frequencies and the predicted probabilities (calibration curves). A calibration curve can be approximated by a regression line with intercept α and slope β. These parameters can be estimated in an LRM with the event as outcome and the linear predictor as the only covariate. Well-calibrated models have $\alpha = 0$ and $\beta = 1$. Therefore, a sensible measure of calibration is a likelihood ratio statistic testing the null hypothesis that $\alpha = 0$ and $\beta = 1$. The statistic has a χ^2 distribution with 2 *df* (unreliability [*U*] statistic).[21] Individual predictions were either calculated from nomograms or were obtained from the original data for the scoring system. We also evaluated average (E average [E aver]) and maximal errors (E maximal [E max]) between predictions and observations obtained from a calibration curve. Calibration is not adequate to evaluate scores or tree-based models, which are intended to give a positive or negative result.

FN rate. For nomograms and RP tools, 10% or less cutoff values were considered to define the subgroup of patients with a low predicted probability of metastatic non-SN. For scores, positive and negative predictions were compared with observed issues.

Clinical utility. The main aim of tools is to identify the largest subgroup of patients with a low risk of non-SN involvement. For each model, we studied the number of patients predicted as negative by scores or having a probability of metastatic non-SN $\leq$ 10% predicted by nomograms or RP models.

OLR model. Multiple linear regression analysis was applied to the population of 561 patients to evaluate the optimal performance on our population. Age, histologic primary tumor size, tumor type, histologic grade, lymphovascular space involvement, estrogen/progesterone receptor status, size of SN metastasis, histologic method of SN analysis, the total number of SNs removed, the numbers of positive and negative SNs, the ratio of positive SNs to all removed SNs, and extracapsular extension were input into the multiple linear regression model with a backward section of informative variables based on Akaike's information criterion. We used the bootstrapping technique to obtain relatively unbiased estimates (200 repetitions); this provides an estimate of the average optimism of the AUC when all data are included. The performance of the OLR (considering overoptimism or not) can



Models	A AUC (95% CI)	B AUC (95% CI)
MSKCC nomogram	0.78 (0.76 to 0.81)	0.72 (0.66 to 0.79)
Mayo nomogram	0.74 (0.71 to 0.76)	0.63 (0.57 to 0.70)
Cambridge nomogram	0.73 (0.7 to 0.75)	0.63 (0.57 to 0.69)
Stanford nomogram	0.72 (0.7 to 0.74)	0.73 (0.68 to 0.79)
Tenon score	0.81 (0.79 to 0.83)	0.81 (0.76 to 0.87)
MDA score	0.73 (0.7 to 0.75)	0.67 (0.62 to 0.73)
Score of Saidi	0.65 (0.62 to 0.67)	0.62 (0.57 to 0.67)
RP-ROC	0.68 (0.65 to 0.70)	0.65 (0.59 to 0.71)
CART	0.65 (0.63 to 0.67)	0.60 (0.54 to 0.65)

Fig 1. Receiver operating characteristic (ROC) curves of the Memorial Sloan-Kettering Cancer Center (MSKCC), Mayo, Cambridge, and Stanford nomograms; Tenon score; The University of Texas M. D. Anderson Cancer Center (MDA) score; Saidi score; and recursive partitioning (RP) –ROC and boosted Classification and Regression Trees (CART) models (A) for the 561 patients and (B) for the 246 patients with micrometastasis or isolated tumor cells in the sentinel lymph node.

be regarded as the maximum that can be expected for a model based on external data.

Evaluation of a subgroup of micrometastasis and ITC. We also tested the accuracy of each model in the subgroup of patients with only micrometastasis or ITC in the SN. Models usually provide poor results for this particular subgroup.

All analyses were performed using the R package with the Design, Hmisc, Rpart, and Verification libraries (http://lib.stat.cmu.edu/R/CRAN/).

RESULTS

Among the 561 patients, at least one non-SN was metastatic in 147 patients (26.2%). We studied the performance of the nine models in terms of discrimination (AUC), calibration, FN rate, and clinical utility (number of patients at low risk for non-SN). We first studied the whole population and then the subgroup of patients with only SN micrometastases or ITC in the SN.

Performance of the Models Over the Whole Population

Discrimination. ROC curves are plotted in Figure 1. When applied to all 561 patients, the model with the highest AUC was the Tenon score with an AUC of 0.81 (95% CI, 0.79 to 0.83). The other models performed as follows: the MSKCC, Mayo, Cambridge, Stanford nomograms, MDA score, Saidi score, RP-ROC, and CART gave AUCs of 0.78 (95% CI, 0.76 to 0.81), 0.73 (95% CI, 0.71 to 0.76), 0.73 (95% CI, 0.70 to 0.75), 0.72 (95% CI, 0.70 to 0.74), 0.73 (95% CI, 0.70 to 0.75), 0.65 (95% CI, 0.62 to 0.67), 0.68 (95% CI, 0.65 to 0.70), and 0.65 (95% CI, 0.63 to 0.67), respectively (Table 3). An OLR model to predict non-SN metastasis was developed from the data of this cohort. The AUC of the resulting model was 0.85 (95% CI, 0.83 to 0.88). The AUC, corrected by bootstrapping, was 0.84, suggesting that an AUC of more than 0.84 is not achievable in this cohort.

Calibration. Calibration plots are given in Figure 2. Two of the four nomograms were well calibrated, with no significant difference between the predicted and the observed probability: Cambridge ($P = .1$) and Mayo ($P = .08$) nomograms. This means that the percentages predicted with other scoring methods were unsatisfactory when both low- and high-risk patients of this series were studied. The average difference (E aver) in predicted and calibrated probabilities ranged from 3% to 23%. The maximal difference (E max) ranged from 5% to 58%. With the OLR model, the *P* value of the *U* index was 1 and the average difference in predicted and calibrated probabilities was 1.6%. The *P* value obtained with the OLR model was as expected; the average difference provides an adequate estimation of the target error when using external models (Table 3).

Table 3. Prospective Validation of the Models for the 561 Positive-SN Patients With Breast Cancer and for the 246 Patients With Micrometastasis or ITC in the SN

Model	Threshold	No. of Patients*	%	AUC	95% CI	Calibration Plot: *P*	E max	E aver	FN Rate %	FN Rate 95% CI	No. of FN Rate†	Range
Overall population, N = 561												
MSKCC nomogram[12]	≤ 10%	201	35.8	0.78	0.76 to 0.81	$< 10^{-3}$	15	6	6.5	3.9 to 10.3	13	7.8-20.7
Mayo nomogram[11]	≤ 10%	16	2.9	0.74	0.71 to 0.76	.08	5	5	12.5	3.5 to 35.6	2	0.6-5.7
Cambridge nomogram[15]	≤ 10%	126	22.5	0.73	0.7 to 0.75	.1	6	3	10.3	6.2 to 12.8	13	7.8-16.1
Stanford nomogram[14]	≤ 10%	101	18	0.72	0.7 to 0.74	$< 10^{-3}$	58	23	4.9	2.2 to 10.7	5	2.2-10.8
Tenon score[10]	≤ 3.5	273	48.7	0.81	0.79 to 0.83	Not adequate			4.4	2.6 to 7.1	12	7.1-19.4
MDA score[6]	≤ 0	140	25	0.73	0.7 to 0.75	Not adequate			5.7	3 to 10.5	8	4.2-14.7
Saidi score[16]	≤ 2	140	25	0.65	0.62 to 0.67	Not adequate			15	10.3 to 21	21	14.4-29.4
RP-ROC[14]	≤ 10	173	30.8	0.68	0.65 to 0.7	Not adequate			5.2	2.8 to 9.2	9	4.8-15.9
CART[14]	≤ 10	186	33.2	0.65	0.63 to 0.67	Not adequate			6.4	3.8 to 10.5	12	7.1-19.5
OLR				0.84, 0.81‡		1	3	1.6				
Subgroup of patients with micrometastasis or ITC in the SN (n = 246)												
MSKCC nomogram[12]	≤ 10	143	58.1	0.72	0.66 to 0.79	.1	6	4	3.5	1.6 to 6.4	5	2.3-9.2
Mayo nomogram[11]	≤ 10	10	4.1	0.63	0.57 to 0.7	$< 10^{-3}$	51	16	10	1.8 to 38	1	0.2-3.8
Cambridge nomogram[15]	≤ 10	73	29.7	0.63	0.57 to 0.69	$< 10^{-3}$	5	10	4.1	1.4 to 9.6	3	1-7
Stanford nomogram[14]	≤ 10%	91	37	0.73	0.68 to 0.79	$< 10^{-3}$	67	20	2.2	0.6 to 6.4	2	0.5-5.8
Tenon score[10]	≤ 3.5	209	85	0.81	0.76 to 0.87	Not adequate			2.9	1.5 to 4.8	6	3.1-10
MDA score[6]	≤ 0	132	53.7	0.67	0.62 to 0.73	Not adequate			3.8	1.7-7	5	2.2-9.2
Saidi score[16]	≤ 2	71	28.9	0.62	0.57 to 0.67	Not adequate			2.8	0.8 to 8.2	2	0.6-5.8
RP-ROC[14]	≤ 10%	173	70.3	0.65	0.59 to 0.71	Not adequate			5.2	3.1 to 7.5	9	5.3-13
CART[14]	≤ 10%	186	75.6	0.6	0.54 to 0.65	Not adequate			6.4	4.3 to 8.3	9	8-15.4

Abbreviations: SN, sentinel lymph node; ITC, isolated tumor cells; AUC, area under the receiver operating characteristic curve; E, difference in predicted and calibrated probabilities between calibration and AUC; E max, maximal error; E aver, average error; FN, false negative; MSKCC, Memorial Sloan-Kettering Cancer Center; MDA, The University of Texas M. D. Cancer Center; RP-ROC, recursive partitioning with receiver operating characteristic; CART, boosted Classification and Regression Trees; OLR, optimal logistic regression model.

*No. of patients with score or predicted probability equal or lower than the threshold.

†Estimation of the no. of patients according to the 95% CI of the FN rate.

‡ORL by bootstrapping.

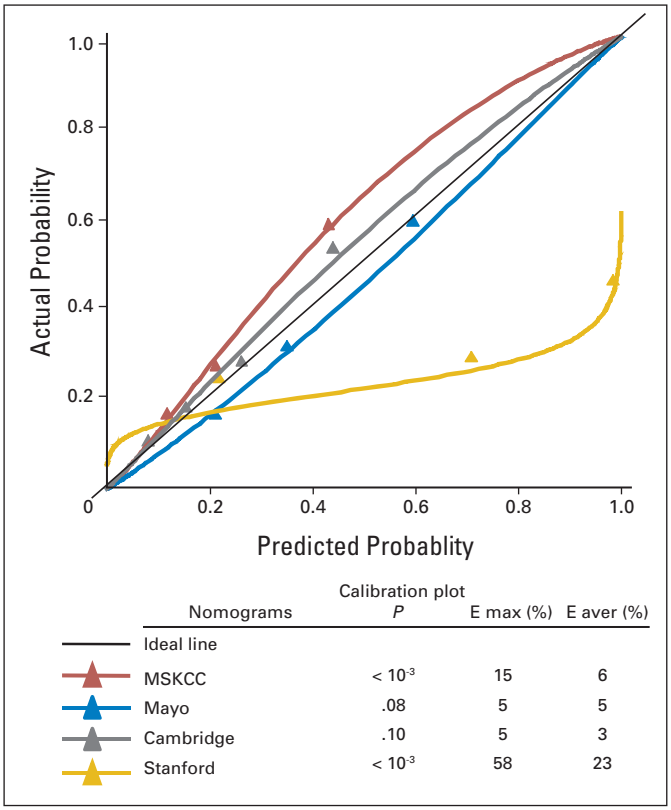


Fig 2. Calibration plot of the Memorial Sloan-Kettering Cancer Center (MSKCC), Mayo, Cambridge, and Stanford nomograms for the entire cohort of 561 patients. E, difference in predicted and calibrated probabilities between calibration and area under the receiver operating characteristic curve; E max, maximal error; E aver, average error.

FN rate. Only two models have an FN rate less than 5%: 4.9% (five of 101) for the Stanford nomogram and 4.4% (12 of 273) for the Tenon score. The MSKCC nomogram, Stanford nomogram, MDA score, RP-ROC, and CART showed FN rates ranging between 5% and 10%. It was greater than 10% for Mayo and Cambridge nomograms and for the Saidi score (Table 3).

Clinical utility. Because the aim of models is to identify the largest possible low-risk subgroup, we specifically studied this aspect. Two RP models and one score assigned more than one third of patients to the low-risk group: RP-ROC, 30.8%; CART, 33.2%; Tenon score, 48%. Except for the MSKCC nomogram (35.8%), other nomograms assigned most patients to the high-risk group (Table 3).

Subgroup of Patients With Only SN Micrometastases or ITC

ROC curves are plotted in Figure 1. The Tenon score outperformed other models with an AUC of 0.81 (95% CI, 0.76 to 0.87), an FN rate of 2.9% (six of 209), and 85% of patients assigned to the low-risk group. Two models had AUCs ranging from 0.7 to 0.8: the MSKCC and Stanford nomograms with AUCs of 0.72 (95% CI, 0.66 to 0.79) and 0.73 (95% CI, 0.68 to 0.79), respectively; these nomograms gave FN rates of 3.5% (five of 143) and 2.2% (two of 91), respectively. Only the MSKCC nomogram was well calibrated ($P = .1$). The six other models gave AUCs lower than 0.7, with the Mayo nomogram showing a 10% higher FN rate. Five models assigned more than half the patients to the low-risk group for non-SN: MSKCC (58.1% of patients assigned), Tenon score (85%), MDA score (53.7%), RP-ROC (70.3%), and CART (75.6%). In contrast, the Mayo nomogram assigned only 4.1% of patients to this group (Table 3).

DISCUSSION

This multicenter study evaluates and compares the performance of nine models currently available to predict non-SN status in patients with breast cancer with SN metastasis; the models were all applied to an independent population of patients with cancer.[6,10-12,14-16] Our study suggests that the different models do not perform equally well, especially for the subgroup of patients with only micrometastasis or ITC in the SN. We indicate evaluation methods that are available to assess model performance and guide clinical practice.

If many models have been developed, few validations have been reported, as shown in Table 4.[11,12,14,15,17,22-33] We believe that our comparison of models and our use of a reasonable target (optimal regression model) will help clinicians to decide what scoring method to use.

Predictive factors of metastatic involvement of non-SLN can be separated into two categories: (1) breast tumors characteristics, and (2) metastatic sentinel lymph node characteristics.[34] Consequently, most of models used tumor size, and number, and burden of metastatic SN as variables. To these robust factors, some models add other variables: lymphovascular invasion could be interesting because for most authors, it indicates the lymphatic spreading of the tumor; however, its input in the model performance is not obvious according to our study: six models integrated this component without improving the results compared with other models. The ratio between positive and total number of nodes is probably more interesting because it combines the number of metastatic SNs and a surrogate for the exhaustiveness of axillary sampling. Some investigators have reported that a sampling of four nodes may provide adequate staging[35,36] or may add information to SN biopsy.[37]

We tested all models with respect to discrimination, calibration, FN rate, and patients assigned to the low-risk group. We found that different evaluations may point out different intrinsic abilities of models. For example, LRM outperformed scores in terms of calibration; this is an expected result because LRM provides continuous probabilities, whereas scores are designed to identify a low-risk group while neglecting individual predictions in the high-risk group.

Discrimination is a popular evaluation criterion. The AUC indicates whether the relative ranking of individual predictions is in the correct order. It does not reflect the accuracy of a model, and its clinical significance is poor. In contrast, the clinical significance of calibration is high: it reflects the accuracy of individual predictions. It is worth emphasizing that of the four nomograms, only Van Zee et al[12] provided calibration measurements. In our study, we calculated average (E aver) and maximal errors (E max) between predictions and observations obtained from calibration curve. This gives an idea of model performance when extrapolated to new patient populations.

Table 4. Validation of Models to Predict Nonsentinel Node Status in Breast Cancer Patients With Metastatic Sentinel Node: Review of Literature

Study	Year	Single- or Multicenter Study	No. of Positive-SN Patients	Area Under the ROC Curve								
				MSKCC Nomogram	Mayo Nomogram	Cambridge Nomogram	Stanford Nomogram	Tenon Score	MDA Score	Saidi Score	RP-ROC	CART
Van Zee et al[12]	2003	Single	373	0.76								
Kocsis et al[27]	2004	Single	140	Not valid								
Soni et al[31]	2005	Single	149	0.75								
Degnim et al[11]	2005											
Mayo clinics dataset		Single	462	0.72	0.77							
Michigan dataset		Single	89	0.86								
Smidt et al[30]	2005	Single	222	0.71								
Specht et al[32]	2005	Single	33	0.72								
Lambert et al[28]	2005	Single	200	0.71								
Cripe et al[24]	2006	Single	92	0.82								
Dauphine et al[25]	2007	Single	39	0.63				0.68	0.7			
Alran et al[22]	2007	Single	588	0.72								
		Single	213*	Not valid								
Ponzone et al[29]	2007	Single	186	0.71					Not valid			
Bevilacqua et al[23]	2007	Single	1545	0,75								
Zgajnar et al[33]	2007	Single	276	0.72								
Pal et al[15]	2008	Single	118	0.68								
Coutant et al[17]	2008	Multi	226					0.82				
Klar et al[26]	2008	Single	98	0.58								
Kohrt et al[14]	2008	Multi	171	0.77			0.85				0.8	0.8
Present study	2008	Multi	561	0.78	0.74	0.73	0.72	0.81	0.73	0.65	0.68	0.65
			246*	0.72	0.63	0.63	0.73	0.81	0.67	0.62	0.65	0.60

Abbreviations: ROC, receiver operating characteristic; SN, sentinel lymph node; MSKCC, Memorial Sloan-Kettering Cancer Center; MDA, The University of Texas M. D. Anderson Cancer Center; RP-ROC, recursive partitioning with receiver operating characteristic; CART, boosted Classification and Regression Trees.
*Subgroup of patients with only SN micrometastasis.

This is of particular importance for clinical practice because probabilities are announced to patients without a CI. To indicate a probability, ±E max or E aver is more appropriate than providing only a probability, for example, "The risk of nonsentinel node metastases is theoretically 10% but it may vary between 4% and 16%."

In fact, nomogram and scores can be used together to inform patients: a negative/positive result is very informative for clinical practice, whereas a quantified prediction is useful for risk assessment. Moreover, the decision for ALND is multivariate, based on patient age, comorbidities, implications for systemic treatment, and not on a model only. In our institutions, 3% of patients with metastatic SLN did not undergo ALND during the study period. This is unlikely that this low number could alter our results. In contrast, in a study published by Dauphine et al, [25] micrometastasis were under-represented because nine of the 14 patients with micrometastasis did not undergo ALND.

Apart from calibration and discrimination, two important criteria of model performance may help clinicians to decide what model to use: a model's FN rate and its ability to identify the largest subgroup of patients with a low risk of non-SN involvement. Nevertheless, the fact remains that no tools have been able to identify patients without any risk of non-SN metastasis. In this study, only two models—Tenon score and Stanford nomogram—gave an FN rate less than 5% (4.4% and 4.9%, respectively), whereas a third of the tools had an FN rate higher than 10% (Mayo nomogram, Cambridge nomogram, and the Saidi score). An FN rate $< 5\%$ is often considered as a target value because this is the false-negative rate of ALND.[10,12,17,25,29] The low proportions of patients for whom ALND could be avoided clearly undermine the reliability of such scores and nomograms for routine clinical practice. For this criterion, the Tenon score outperformed other models, with 48% of patients identified at low risk for non-SN metastasis. Only three others models—MSKCC nomogram and the two RP models—assigned more than 30% of patients to the group for whom ALND could be avoided (Table 3), whereas the other models assigned low proportions of patients to the low-risk group. Interestingly, these latter models also showed low clinical utility with their original patient cohorts, suggesting that our findings are not specific to the patient cohort in the present analysis.

The subgroup of patients with only micrometastases or ITC is investigated in studies like this one because most of these patients were negative for non-SN.[38] In theory, therefore, a model could perform better for this subset of patients than for the whole population. We found here that 44% of patients had only micrometastases or ITC in SN. This proportion was comparable to those reported by other studies.[6,26,29] Alran et al[22] tested the accuracy of the MSKCC nomogram in 588 patients and found that it was reliable for patients with macrometastatic SNs but not for those with micrometastatic SNs.[14,22] To our knowledge, no other model has been evaluated for this specific subgroup.

In this study, we showed that the Tenon score is particularly accurate for this subgroup of patients, with an AUC of 0.81 (95% CI, 0.76 to 0.87) and an FN rate of 2.9% (six of 209). Only two other

models presented an AUC higher than 0.7: the MSKCC nomogram and Stanford nomogram, with AUCs of 0.72 (95% CI, 0.66 to 0.79) and 0.73 (95% CI, 0.68 to 0.79), respectively, and FN rates of 3.5% (five of 143) and 2.2% (two of 91), respectively (Table 3). In contrast to Alran et al,[22] we validated the MSKCC nomogram for this subgroup of patients. In their study, the MSKCC nomogram was used for positive SN biopsy findings regardless of the method of metastasis detection, and some patients had SN detection by blue dye only. This may explain why Alran et al[22] failed to validate the MSKCC nomogram in the subgroup of patients with micrometastatic SNs.

AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST

The author(s) indicated no potential conflicts of interest.

AUTHOR CONTRIBUTIONS

Conception and design: Charles Coutant, Emmanuel Barranger, Serge Uzan, Roman Rouzier
Administrative support: Charles Coutant, Roman Rouzier
Provision of study materials or patients: Charles Coutant, Eric Lambaudie, Eric Fondrinier, Fréderic Marchal, François Guillemin, Nathalie Seince, Véronique Thomas, Jean Levêque, Emmanuel Barranger, Emile Darai, Gilles Houvenaeghel
Collection and assembly of data: Charles Coutant, Camille Olivier
Data analysis and interpretation: Charles Coutant, Camille Olivier, Roman Rouzier
Manuscript writing: Charles Coutant, Roman Rouzier
Final approval of manuscript: Charles Coutant, Camille Olivier, Eric Lambaudie, Eric Fondrinier, Fréderic Marchal, François Guillemin, Nathalie Seince, Véronique Thomas, Jean Levêque, Emmanuel Barranger, Emile Darai, Serge Uzan, Gilles Houvenaeghel, Roman Rouzier

REFERENCES

1. Lyman GH, Giuliano AE, Somerfield MR, et al: American Society of Clinical Oncology guideline recommendations for sentinel lymph node biopsy in early-stage breast cancer. J Clin Oncol 23:7703-7720, 2005
2. Schwartz GF, Giuliano AE, Veronesi U: Proceedings of the consensus conference on the role of sentinel lymph node biopsy in carcinoma of the breast, April 19-22, 2001, Philadelphia, Pennsylvania. Cancer 94:2542-2551, 2002
3. Veronesi U, Paganelli G, Galimberti V, et al: Sentinel-node biopsy to avoid axillary dissection in breast cancer with clinically negative lymph-nodes. Lancet 349:1864-1867, 1997
4. Chu KU, Turner RR, Hansen NM, et al: Do all patients with sentinel node metastasis from breast carcinoma need complete axillary node dissection? Ann Surg 229:536-541, 1999
5. Chu KU, Turner RR, Hansen NM, et al: Sentinel node metastasis in patients with breast carcinoma accurately predicts immunohistochemically detectable nonsentinel node metastasis. Ann Surg Oncol 6:756-761, 1999
6. Hwang RF, Krishnamurthy S, Hunt KK, et al: Clinicopathologic factors predicting involvement of nonsentinel axillary nodes in women with breast cancer. Ann Surg Oncol 10:248-254, 2003
7. Kamath VJ, Giuliano R, Dauway EL, et al: Characteristics of the sentinel lymph node in breast cancer predict further involvement of higher-echelon nodes in the axilla: A study to evaluate the need for complete axillary lymph node dissection. Arch Surg 136:688-692, 2001
8. Kim T, Giuliano AE, Lyman GH: Lymphatic mapping and sentinel lymph node biopsy in early-stage breast carcinoma: A metaanalysis. Cancer 106:4-16, 2006
9. Nos C, Harding-MacKean C, Freneaux P, et al: Prediction of tumour involvement in remaining axillary lymph nodes when the sentinel node in a woman with breast cancer contains metastases. Br J Surg 90:1354-1360, 2003
10. Barranger E, Coutant C, Flahault A, et al: An axilla scoring system to predict non-sentinel lymph node status in breast cancer patients with sentinel lymph node involvement. Breast Cancer Res Treat 91:113-119, 2005
11. Degnim AC, Reynolds C, Pantvaidya G, et al: Nonsentinel node metastasis in breast cancer patients: Assessment of an existing and a new predictive nomogram. Am J Surg 190:543-550, 2005
12. Van Zee KJ, Manasseh DM, Bevilacqua JL, et al: A nomogram for predicting the likelihood of additional nodal metastases in breast cancer patients with a positive sentinel node biopsy. Ann Surg Oncol 10:1140-1151, 2003
13. Viale G, Maiorano E, Pruneri G, et al: Predicting the risk for additional axillary metastases in patients with breast carcinoma and positive sentinel lymph node biopsy. Ann Surg 241:319-325, 2005
14. Kohrt HE, Olshen RA, Bermas HR, et al: New models and online calculator for predicting non-sentinel lymph node status in sentinel lymph node positive breast cancer patients. BMC Cancer 8:66, 2008
15. Pal A, Provenzano E, Duffy SW, et al: A model for predicting non-sentinel lymph node metastatic disease when the sentinel lymph node is positive. Br J Surg 95:302-309, 2008
16. Saidi RF, Dudrick PS, Remine SG, et al: Non-sentinel lymph node status after positive sentinel lymph node biopsy in early breast cancer. Am Surg 70:101-105, 2004; discussion 105
17. Coutant C, Rouzier R, Fondrinier E, et al: Validation of the Tenon breast cancer score for predicting non-sentinel lymph node status in breast cancer patients with sentinel lymph node metastasis: A prospective multicenter study. Breast Cancer Res Treat 2008
18. Salford Systems: CART software. http://www.salford-systems.com/cart.php
19. Kraemer H: Evaluating medical tests: Objective and quantitative guidelines. Newbury Park, CA, Sage Publications, 1992
20. Hanley JA, McNeil BJ: The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 143:29-36, 1982
21. Cox D: Two further applications of a model for binary regression. Biometrika 45:562-565, 1958
22. Alran S, De Rycke Y, Fourchotte V, et al: Validation and limitations of use of a breast cancer nomogram predicting the likelihood of non-sentinel node involvement after positive sentinel node biopsy. Ann Surg Oncol 14:2195-2201, 2007
23. Bevilacqua JL, Kattan MW, Fey JV, et al: Doctor, what are my chances of having a positive sentinel node? A validated nomogram for risk estimation. J Clin Oncol 25:3670-3679, 2007
24. Cripe MH, Beran LC, Liang WC, et al: The likelihood of additional nodal disease following a positive sentinel lymph node biopsy in breast cancer patients: Validation of a nomogram. Am J Surg 192:484-487, 2006
25. Dauphine CE, Haukoos JS, Vargas MP, et al: Evaluation of three scoring systems predicting non sentinel node metastasis in breast cancer patients with a positive sentinel node biopsy. Ann Surg Oncol 14:1014-1019, 2007
26. Klar M, Jochmann A, Foeldi M, et al: The MSKCC nomogram for prediction the likelihood of non-sentinel node involvement in a German breast cancer population. Breast Cancer Res Treat 112: 523-531, 2008
27. Kocsis L, Svebis M, Boross G, et al: Use and limitations of a nomogram predicting the likelihood of non-sentinel node involvement after a positive sentinel node biopsy in breast cancer patients. Am Surg 70:1019-1024, 2004
28. Lambert LA, Ayers GD, Hwang RF, et al: Validation of a breast cancer nomogram for predicting nonsentinel lymph node metastases after a positive sentinel node biopsy. Ann Surg Oncol 13: 310-320, 2006
29. Ponzone R, Maggiorotto F, Mariani L, et al: Comparison of two models for the prediction of nonsentinel node metastases in breast cancer. Am J Surg 193:686-692, 2007
30. Smidt ML, Kuster DM, van der Wilt GJ, et al: Can the Memorial Sloan-Kettering Cancer Center nomogram predict the likelihood of nonsentinel lymph node metastases in breast cancer patients in the Netherlands? Ann Surg Oncol 12:1066-1072, 2005
31. Soni NK, Carmalt HL, Gillett DJ, et al: Evaluation of a breast cancer nomogram for prediction of non-sentinel lymph node positivity. Eur J Surg Oncol 31:958-964, 2005
32. Specht MC, Kattan MW, Gonen M, et al: Predicting nonsentinel node status after positive sentinel lymph biopsy for breast cancer: Clinicians versus nomogram. Ann Surg Oncol 12:654-659, 2005
33. Zgajnar J, Perhavec A, Hocevar M, et al: Low performance of the MSKCC nomogram in preoperatively ultrasonically negative axillary lymph node in breast cancer patients. J Surg Oncol 96:547-553, 2007
34. Travagli JP, Atallah D, Mathieu MC, et al: Sentinel lymphadenectomy without systematic axillary dissection in breast cancer patients: Predictors

of non-sentinel lymph node metastasis. Eur J Surg Oncol 29:403-406, 2003

35. Chetty U, Jack W, Prescott RJ, et al: Management of the axilla in operable breast cancer treated by breast conservation: A randomized clinical trial. Edinburgh Breast Unit. Br J Surg 87:163-169, 2000

36. Tanaka K, Yamamoto D, Kanematsu S, et al: A four node axillary sampling trial on breast cancer patients. Breast 15:203-209, 2006

37. Adwani A, Ebbs SR, Burton S, et al: Sentinel node biopsy should be supplemented by axillary sampling in patients with small breast cancers. Int Semin Surg Oncol 2:27, 2005

38. Houvenaeghel G, Nos C, Mignotte H, et al: Micrometastases in sentinel lymph node in a multicentric study: Predictive factors of nonsentinel lymph node involvement—Groupe des Chirurgiens de la Federation des Centres de Lutte Contre le Cancer. J Clin Oncol 24:1814-1822, 2006

Acknowledgment

Charles Coutant thanks l'Association pour la Recherche sur le Cancer, le Ministère Français des Affaires Etrangères et Européennes as part of the Lavoisier program, le Collège National des Gynécologues et Obstétriciens Français, and la Fondation de France as part of its collaboration with la Fedération Nationale des Centres de Lutte Contre le Cancer.

