Published in final edited form as:

Ann Surg Oncol. 2010 November ; 17(11): 2884–2891. doi:10.1245/s10434-010-1077-1.

Predicting Four or More Metastatic Axillary Lymph Nodes in Sentinel Node Positive Breast Cancer Patients: Assessment of Existent Risk Scores

Benjamin Zendejas, MD¹, Tanya L. Hoskin, MS², Amy C. Degnim, MD¹, Carol A. Reynolds, MD³, David R. Farley, MD¹, and Judy C. Boughey, MD¹

¹Department of Surgery, Mayo Clinic, Rochester, MN, USA

²Department of Health Sciences Research, Mayo Clinic, Rochester, MN, USA

³Department of Pathology, Mayo Clinic, Rochester, MN, USA

Abstract

Background—Patients with metastases in four or more axillary lymph nodes (4+ALN) represent a subset of breast cancer patients at increased risk of local recurrence and who benefit from post-mastectomy radiation. Risk prediction models designed to identify such patients have been published by Rivers et al, Chagpar et al, and Katz et al. We sought to evaluate and compare the performance of these models in an independent patient population.

Methods—We reviewed 454 breast cancer patients with 1–3 positive sentinel lymph nodes who underwent completion axillary lymph node dissection at our institution. Each of the 3 published models was applied to our sample as described in the respective publications. The models' performances were analyzed with the Hosmer-Lemeshow goodness-of-fit test and with the area under the curve (AUC). Sensitivity, specificity, and false negative percentages were calculated for clinically meaningful cut-off points of each score.

Results—Of 454 eligible patients, 87 (19.2%) had 4 positive axillary nodes. The Rivers, Chagpar and Katz models demonstrated good calibration in our population based on the Hosmer-Lemeshow test ($p=0.82$, $p=0.73$, $p=0.71$, respectively). Assessment of discriminatory ability for the models resulted in AUCs of 0.81, 0.73 and 0.81, respectively.

Conclusions—The Rivers and Katz models performed well in our patient population and may be clinically useful to predict patients with 4+ALN. However, their clinical utility is limited by the current controversy surrounding the use of post-mastectomy radiation for all node positive patients.

Keywords

Breast Cancer; nomogram; prediction model; sentinel lymph node metastases; post-mastectomy radiation therapy

Introduction

Axillary node status is the most important prognostic factor in breast cancer¹. Sentinel lymph node (SLN) surgery has emerged as the preferred technique for axillary staging. With a low morbidity and high negative predictive value², axillary lymph node dissection

Address correspondence and reprint request to: Judy C. Boughey, MD, Mayo Clinic 200 First Street SW, Rochester MN, 55905, USA, boughey.judy@mayo.edu.

(ALND) can be safely omitted in patients with a negative SLN. On the other hand, a completion axillary lymph node dissection (CALND) is standard of care for patients with a positive SLN. However, this practice has been questioned because 40% to 60% of patients have no disease in the remaining axillary lymph nodes (ALNs)^{3, 4}, and for these patients CALND is unlikely to be beneficial. In addition, randomized trials with long term follow up suggest that prophylactic resection of occult axillary metastases is comparable to observation and delayed therapeutic ALND for cases of regional recurrence^{5, 6}.

Post mastectomy radiation therapy (PMRT) or extended field radiation after lumpectomy is considered standard of care for patients with four or more positive axillary lymph nodes (4+ALNs)⁷. These high risk patients, regardless of the size of their primary tumor, are at increased risk for local and regional recurrence and therefore more extensive locoregional therapy is recommended. However, for patients with one to three positive SLNs, it is not known whether they will have 4+ALNs on final pathology after CALND, in which case nodal radiation would be indicated. Even though some series have reported that immediate breast reconstruction in the setting of PMRT can be done with good cancer control and acceptable cosmesis,^{8, 9} most surgeons and radiation oncologists would prefer to delay reconstruction if PMRT is anticipated.¹⁰ Thus, although controversy exists in whether or not to proceed with immediate breast reconstruction, being able to predict who will have 4+ALNs on final pathology can substantially benefit the decision making process of those who prefer delayed reconstruction in the setting of PMRT.

These conundrums have led to the development of statistical models to predict how likely a patient is to harbor additional metastatic nodes after resection of at least one positive SLN. The goal of these tools is to refine the selection of SLN positive patients who will benefit from CALND due to additional metastases in non-SLNs^{11–13}. In addition, another set of prediction models have been developed to identify high risk SLN positive patients who will likely harbor 4+ALNs^{14–16}.

Three distinct models have been developed to identify patients likely to harbor 4+ALNs. The first was published by Rivers et al, from the University of Michigan and M. D. Anderson Cancer Center¹⁵. The Rivers model assigns predicted probabilities of having 4+ALN depending on variables including tumor size, number of positive SLNs, percent of positive SLNs, and presence or absence of lymphovascular invasion and/or extranodal extension, respectively. Subsequently, Chagpar et al¹⁴ introduced a scoring system based on available pre-operative and intra-operative variables. This model assigns one point for each positive SLN, one point if the proportion of SLNs positive was 50% or greater, one point if the tumor was T2 instead of T1, and one point if the SLN metastasis was found using hematoxylin-eosin staining (as opposed to immunohistochemistry), providing a total maximum score of six points used to yield a corresponding predicted probability. Most recently, to delineate the extent of radiation therapy for patients with a positive SLN and do not undergo CALND, Katz et al¹⁶ developed a nomogram to identify patients likely to have 4+ALNs. This model incorporates primary tumor size, number of involved SLNs, size of largest SLN metastasis, lymphovascular invasion, extranodal extension, histology and number of negative SLNs.

To our knowledge such risk prediction models to identify patients likely to harbor 4+ALNs have not been evaluated and compared in an independent population. The objectives of this study were to assess and compare the performance of these scoring systems by applying them to our data set of SLN positive patients with CALND.

Materials and Methods

With the approval of the Mayo Clinic Institutional Review Board, medical records were reviewed of 465 clinically lymph node-negative breast cancer patients with 1–3 positive SLN(s) who underwent CALND at our institution from October 1997 through June 2004. The inclusion criteria, method of SLN identification, pathologic review, tumor and SLN features are described in previously published reports^{13, 17, 18}. The performance of several models^{11,12, 17} for the prediction of additional non SLN disease has been evaluated in this population, and the Mayo model was developed using this cohort¹³. Briefly, SLNs were identified using radioisotope and/or blue dye according to surgeon preference. Intraoperative frozen section and permanent hematoxylin and eosin (H&E) stains were performed on all SLNs. Immunohistochemical (IHC) stains were performed on all SLNs that were negative by frozen section examination and permanent H&E. SLNs were considered positive if metastasis was noted on IHC or H&E. The database also includes variables for age, pathologic size of primary invasive breast tumor, histology, lymphovascular invasion, extranodal extension, size of largest SLN metastasis, number of positive and negative SLNs, and number of positive non SLNs for each patient (Table 1).

Each of the three published models was applied to our sample as described in the respective publications. The Chagpar model could only be applied to patients with tumor stages 1 or 2 based on the sample used to create that model and the model definition (i.e., one point added for being T2 vs T1); thus, 39 patients with T3 tumors in our sample did not have Chagpar scores calculated, but these patients are included in the evaluation of the Rivers and Katz models. The Chagpar model also added one point to the score if the SLN metastasis was detected by H&E instead of IHC; our patients had frozen section as an additional option for the method of detection; since the model did not allow for this possibility, we treated this as H&E detection in applying the model. The scores derived from applying the logistic regression coefficients reported in Table 3 of the Rivers publication were converted to predicted probabilities in the standard fashion for a logistic regression equation: $\text{predicted probability} = \exp(\text{score}) / (1 + \exp(\text{score}))$.

The Rivers model results in a score that is interpretable as the predicted probability of 4+ALNs, and as such ranges from 0 to 1. The Chagpar model results in an integer score from 1–6 with higher scores corresponding to higher likelihood of 4+ALNs. The Katz model results in a score that will range between –53 and +45 in most patients; negative scores translate to probabilities of 4+ALNs that are < 0.50, while positive scores translate to probabilities > 0.50. Figure 2 from the Katz et al paper summarizes how the total calculated score can be converted to likelihood of 4+ALNs for several possible scores.

Model discrimination was assessed by comparing average scores between those with four or more versus less than four positive nodes using a two-sample t-test. The receiver operating characteristic (ROC) curve was generated for each model and the area under the curve (AUC) with 95% confidence intervals was calculated. AUCs were compared across models using the DeLong error method. Model calibration was assessed by the Hosmer-Lemeshow goodness-of-fit statistic and by comparing observed proportions and their 95% exact binomial confidence intervals in specific score ranges to the model-based or reported probabilities in those ranges; p-values were calculated for such comparisons using the binomial distribution. In addition, loess-estimated curves of the relationship between predicted and actual probability were plotted for a visual assessment of calibration. Sensitivity, specificity, and false negative percentages were reported for clinically meaningful cut-off points of each score. Sensitivity and specificity estimates were compared between prediction rules using McNemar's test for paired proportions or the analogous sign test in cases where the number of discordant pairs was < 20. Analysis was performed using

SAS (Version 9.1.3, SAS Institute Inc., Cary, NC) and R (Version 2.7.1, R Foundation for Statistical Computing, Vienna, Austria). Calibration figure was created using the R package Design.¹⁹ P-values < 0.05 were considered statistically significant.

Results

Out of the 465 eligible patients in our database, 10 patients were excluded because they had more than four positive SLNs. Of the remaining 455 patients, one patient with unknown tumor size was excluded because tumor size is necessary for each of the models being validated. Thus, the analysis sample includes 454 patients with complete data. Of these, 87 (19.2%) had 4+ALNs (positive SLNs + positive non SLNs). Characteristics of the overall cohort and of the patients with 4+ALNs are described in detail in Table 1.

A summary of model discrimination appears in Table 2. Each of the scores was significantly associated with the presence 4+ALNs ($p < 0.0001$). However, the score ranges in the two outcome categories overlap almost entirely for each model. Both the Rivers and Katz models had an AUC value of 0.81 in our sample, while the AUC of the Chagpar model was significantly lower than both of these at 0.73 ($p = 0.01$ for both comparisons). This difference in AUC values is represented in Figure 1.

With respect to calibration, the Hosmer and Lemeshow goodness-of-fit statistics had non-significant p-values (Rivers $p=0.82$, Chagpar $p=0.73$, and Katz $p=0.81$) for scores derived from each of the three models suggesting that none stands out as particularly superior or inferior in this regard. We examined the observed proportions with four or more nodes positive in specific score categories (Table 3) and compared them to the published or model-based probabilities associated with those scores. Generally, each of the models seemed to calibrate fairly well to the reported or model-based estimates of the proportion with four or more positive nodes. Intervals where the observed proportion from our sample fell outside the predicted/reported range are shown in bold. We note, however, that these are more commonly overestimates than underestimates, that they are often in the higher probability regions, and that in no case are these differences large enough that we can say our observed proportion is significantly different from the predicted/reported proportion. The first two points are confirmed by examining the calibration curves for each model (Figure 2). The probabilities reported by Chagpar consistently overestimated the actual probabilities in our sample. On the other hand, the Katz model showed a trend toward underestimating the probability in the lowest two categories of probability, which is of more concern given how these probabilities may be used to make clinical decisions. Those with a score of ≤ -53 or less were predicted to have a likelihood 4+ALNs of 0.005 or less; yet 1/21 (5%) of our patients in this category had 4+ALNs; the probability of observing 1/21 if the true probability were 0.005 is $p = 0.10$. Combining the bottom two categories, we observed 2/48 (4%) with

4+ALNs; the chance of observing this number if the true probability were 0.01 (the upper end of the range) is $p=0.08$. If the Katz nomogram were used with a cutpoint of ≤ -29 for a probability of 0.05 or less, we found 4% with 4+ALNs in that range. Therefore in general the model performs reasonably well at low probabilities, but we have reason to believe that the lowest probabilities (0.005 and 0.01) might be underestimates.

Accuracy statistics at the cut-off points of 5% and 10% predicted probability are reported for the Rivers and Katz models and at the cut-off points of 1, 2, or 3 points for the Chagpar model in Table 4. Since we observed zero patients with a Chagpar score of 1 and only two patients with a Chagpar score of 2 who had 4+ALNs, those cut-off points provide 100% and 97% sensitivity, respectively, but also have very poor specificity at only 1% and 20%, respectively. A Rivers cutpoint of 5% yielded sensitivity of 90% and specificity of 46%,

while the Katz model at the same cutpoint had sensitivity of 93% and specificity of 44% ($p=0.45$, $p=0.28$ for sensitivity and specificity respectively).

Discussion

The prediction and possible identification of high risk patients more likely to harbor 4+ALNs could be of clinical use in the following scenarios: 1) intraoperative decision making regarding the delay of immediate reconstruction if PMRT is anticipated in patients with primary tumors <5cm and one to three positive SLN, and 2) planning the dosing and extent of radiotherapy for patients with positive SLNs who do not undergo CALND.

For a variety of reasons, some patients with positive SLNs do not undergo CALND²⁰. The appropriate treatment for these patients is unclear. In these circumstances, radiation oncologists must decide whether radiotherapy to the regional nodes is warranted with limited information about the extent of axillary involvement. After breast conserving surgery, one option is to perform a standard tangential field approach, extending the borders of the radiation field to adequately treat up to 80% of Level I/II axillary lymph nodes²¹. The issue is that patients with 4+ALNs have a higher risk of locoregional recurrence in the supraclavicular fossa or axillary apex²². To decrease this risk, some radiation oncologists advocate the addition of a supraclavicular/axillary apex field to tangential breast radiation in a more comprehensive approach, but this approach also carries an increase in treatment-related morbidity. Therefore for women who do not undergo CALND, it would be valuable to predict which patients are at low risk for harboring 4+ALNs and could be spared the morbidity of comprehensive nodal radiation.

Risk models designed to predict the need for PMRT can have a clinical use assisting patients and their treating physicians in the complex decision-making process regarding their breast cancer treatment. Patients who would otherwise choose to have a mastectomy in order to try to avoid the adjuvant radiation therapy mandated with breast conservation therapy may choose partial mastectomy if adjuvant radiation is inevitable²³. Unfortunately, such models would need to rely only on preoperative information and are not available yet.

The Rivers Model

When comparing our cohort with the one used to generate the Rivers model, the Mayo Cohort had a higher proportion of patients with 4+ALNs (19% vs 14%) and tended to have more patients in worse categories, such as a higher proportion of patients with tumor size >2cm, lymphovascular invasion and extranodal extension.

To our knowledge this is the first assessment of the performance of the Rivers model in an independent population and although no AUC is reported in the original Rivers et al publication we found that the Rivers model achieved good discrimination in our patient population. Both the Rivers and Katz models use extranodal extension as a predictor variable. One could argue that because some radiation oncologists would offer PMRT for patients with even one positive LN demonstrating extranodal extension,²⁴ these models may not alter clinical management. Nonetheless, the primary value of these models is to serve as an adjunct to clinical judgment; if PMRT will be recommended on the basis of extranodal extension then the model can be used selectively for those in whom extranodal extension is absent.

The Chagpar Model

The performance of the Chagpar model was poor in our sample with a significantly lower AUC (0.73) than the other two models and compared to those reported in the original training and validation samples (0.882 and 0.895)¹⁴. We believe that the method of detection

may have accounted for the lower performance of the model in our study since 81% of our sample was detected by frozen section and their model only allowed for H&E or IHC detection. In addition, the Chagpar model is limited by the fact that it does not take into account T3 patients and does not include other factors that have been extensively shown to be predictive of extensive nodal disease in multivariate analysis, such as lymphovascular invasion and extranodal extension²⁵.

The Katz Model

The performance of the Katz model in the study and test sets of the original publication yielded AUCs of 0.83 and 0.81, very similar to the 0.81 encountered in this study. The Katz model is simpler to use when compared to the Rivers model, but it could still be made easier with an online calculator. Although the Katz model performed fairly well overall, it tended to underestimate at the very low probabilities. For example, if someone is told that the probability is less than 0.005, he/she thinks that at most 5/1000 or 1/200 will have 4+ALNs, but we observed one patient with 4+ALNs in only 21 patients who fell into that score range. Despite this, the group with Katz probability 0.05 seemed on target as a whole with an observed proportion of 0.04 in our sample.

Conclusions

The performance of both the Rivers and Katz models were good in our patient population. These risk prediction tools could be useful for: 1) those who would prefer to delay reconstruction if PMRT is recommended, 2) those who would offer PMRT only for those with 4+ALNs, 3) those who would obviate radiation therapy for patients with positive SLN who do not undergo CALND and have low probability of having 4+ALNs.

It is important to bear in mind that current nomograms do not advise clinicians when to delay immediate reconstruction due to a high likelihood of PMRT or when to proceed with immediate reconstruction due to a low likelihood of PMRT, rather they provide probability estimates to be taken into consideration with the overall clinical picture. In addition, specificity values are not perfect for any of the models, and the ranges almost completely overlap between the two outcome groups for each of the scores, so there may be room for improvement although it is not clear that additional variables would be useful.

The issue of treating all node positive patients with PMRT remains controversial. If the number of the metastatic nodes does not influence the decision for nodal radiation, then risk factors and nomograms for the prediction of 4+ALNs will not be helpful when discussing treatment options with breast cancer patients. Until then, decision making regarding immediate reconstruction and need for PMRT in the setting of one to three positive SLNs as well as the extent of radiation therapy for SLN positive patients who do not undergo CALND must be done in a multidisciplinary manner.

Acknowledgments

This publication was made possible by Grant Number 1 UL1 RR024150 from the National Center for Research Resources (NCRR), a component of the National Institutes of Health (NIH), and the NIH Roadmap for Medical Research. Its contents are solely the responsibility of the authors and do not necessarily represent the official view of NCRR or NIH.

References

1. Fisher B, Bauer M, Wickerham DL, et al. Relation of number of positive axillary nodes to the prognosis of patients with primary breast cancer. An NSABP update. *Cancer*. Nov 1; 1983 52(9): 1551–1557. [PubMed: 6352003]

2. Harlow SP, Krag DN, Julian TB, et al. Prerandomization Surgical Training for the National Surgical Adjuvant Breast and Bowel Project (NSABP) B-32 trial: a randomized phase III clinical trial to compare sentinel node resection to conventional axillary dissection in clinically node-negative breast cancer. *Ann Surg.* Jan; 2005 241(1):48–54. [PubMed: 15621990]
3. Abdessalam SF, Zervos EE, Prasad M, et al. Predictors of positive axillary lymph nodes after sentinel lymph node biopsy in breast cancer. *Am J Surg.* Oct; 2001 182(4):316–320. [PubMed: 11720662]
4. Reynolds C, Mick R, Donohue JH, et al. Sentinel lymph node biopsy with metastasis: can axillary dissection be avoided in some patients with breast cancer? *J Clin Oncol.* Jun; 1999 17(6):1720–1726. [PubMed: 10561208]
5. Fisher B, Jeong JH, Anderson S, Bryant J, Fisher ER, Wolmark N. Twenty-five-year follow-up of a randomized trial comparing radical mastectomy, total mastectomy, and total mastectomy followed by irradiation. *N Engl J Med.* Aug 22; 2002 347(8):567–575. [PubMed: 12192016]
6. Veronesi U, Cascinelli N, Mariani L, et al. Twenty-year follow-up of a randomized study comparing breast-conserving surgery with radical mastectomy for early breast cancer. *N Engl J Med.* Oct 17; 2002 347(16):1227–1232. [PubMed: 12393819]
7. Harris JR, Halpin-Murphy P, McNeese M, Mendenhall NP, Morrow M, Robert NJ. Consensus Statement on postmastectomy radiation therapy. *Int J Radiat Oncol Biol Phys.* Jul 15; 1999 44(5): 989–990. [PubMed: 10421530]
8. Jhaveri JD, Rush SC, Kostroff K, et al. Clinical outcomes of postmastectomy radiation therapy after immediate breast reconstruction. *Int J Radiat Oncol Biol Phys.* Nov 1; 2008 72(3):859–865. [PubMed: 18472362]
9. Cordeiro PG, Pusic AL, Disa JJ, McCormick B, VanZee K. Irradiation after immediate tissue expander/implant breast reconstruction: outcomes, complications, aesthetic results, and satisfaction among 156 patients. *Plast Reconstr Surg.* Mar; 2004 113(3):877–881. [PubMed: 15108879]
10. Kronowitz SJ, Robb GL. Radiation therapy and breast reconstruction: a critical review of the literature. *Plast Reconstr Surg.* Aug; 2009 124(2):395–408. [PubMed: 19644254]
11. 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.* Dec; 2003 10(10):1140–1151. [PubMed: 14654469]
12. 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.* 2008; 8:66. [PubMed: 18315887]
13. 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.* Oct; 2005 190(4): 543–550. [PubMed: 16164917]
14. Chagpar AB, Scoggins CR, Martin RC 2nd, et al. Predicting patients at low probability of requiring postmastectomy radiation therapy. *Ann Surg Oncol.* Feb; 2007 14(2):670–677. [PubMed: 17096055]
15. Rivers AK, Griffith KA, Hunt KK, et al. Clinicopathologic features associated with having four or more metastatic axillary nodes in breast cancer patients with a positive sentinel lymph node. *Ann Surg Oncol.* Jan; 2006 13(1):36–44. [PubMed: 16378156]
16. Katz A, Smith BL, Golshan M, et al. Nomogram for the prediction of having four or more involved nodes for sentinel lymph node-positive breast cancer. *J Clin Oncol.* May 1; 2008 26(13):2093–2098. [PubMed: 18445838]
17. Scow JS, Degnim AC, Hoskin TL, Reynolds C, Boughhey JC. Simple prediction models for breast cancer patients with solitary positive sentinel nodes—are they valid? *Breast J.* Nov-Dec; 2009 15(6):610–614. [PubMed: 19824998]
18. Scow JS, Degnim AC, Hoskin TL, Reynolds C, Boughhey JC. Assessment of the performance of the Stanford Online Calculator for the prediction of nonsentinel lymph node metastasis in sentinel lymph node-positive breast cancer patients. *Cancer.* Sep 15; 2009 115(18):4064–4070. [PubMed: 19517477]
19. Harrel, Frank E, Jr. Design: Design Package R. Version 2.1–2.12. 2007. <http://biostat.mc.vanderbilt.edu/s/Design>

20. Zakaria S, Pantvaidya G, Reynolds CA, et al. Sentinel node positive breast cancer patients who do not undergo axillary dissection: are they different? *Surgery*. May; 2008 143(5):641–647. [PubMed: 18436012]
21. Schlembach PJ, Buchholz TA, Ross MI, et al. Relationship of sentinel and axillary level I–II lymph nodes to tangential fields used in breast irradiation. *Int J Radiat Oncol Biol Phys*. Nov 1; 2001 51(3):671–678. [PubMed: 11597808]
22. Taghian A, Jeong JH, Mamounas E, et al. Patterns of locoregional failure in patients with operable breast cancer treated by mastectomy and adjuvant chemotherapy with or without tamoxifen and without radiotherapy: results from five National Surgical Adjuvant Breast and Bowel Project randomized clinical trials. *J Clin Oncol*. Nov 1; 2004 22(21):4247–4254. [PubMed: 15452182]
23. Chagpar AB. Predicting extensive nodal disease in women with breast cancer. *Ann Surg Oncol*. Jan; 2006 13(1):3–4. [PubMed: 16372148]
24. Fisher BJ, Perera FE, Cooke AL, et al. Extracapsular axillary node extension in patients receiving adjuvant systemic therapy: an indication for radiotherapy? *Int J Radiat Oncol Biol Phys*. Jun 1; 1997 38(3):551–559. [PubMed: 9231679]
25. Degnim AC, Griffith KA, Sabel MS, et al. Clinicopathologic features of metastasis in nonsentinel lymph nodes of breast carcinoma patients. *Cancer*. Dec 1; 2003 98(11):2307–2315. [PubMed: 14635063]

Synopsis

Patients with four or more positive axillary nodes benefit from postmastectomy radiation therapy. We evaluated and compared the performance of three models predicting the risk of four or more positive axillary nodes in an independent patient population.

FIG. 1.

Receiver operating characteristic (ROC) curves for each of the three models predicting presence of 4 positive nodes applied to our validation sample are shown. The area under the curve (AUC) was 0.81 for both the Rivers and Katz models but only 0.73 ($p = 0.01$) for the Chagpar model

FIG. 2.

Loess calibration curves for each model relative to the line of identity or ideal calibration (intercept = 0, slope = 1). The maximum and average absolute calibration error (difference between Loess-estimated calibration curve and the line of identity) are reported for each model as $E_{\max}$ and E_{avg} , respectively

Table 1

Demographic and clinical characteristics of the 454 breast cancer cases with 1–3 positive SLNs and CALND that comprised our validation sample are summarized overall and separately according to whether the total number of positive axillary nodes was 4 or more versus less than 4.

	Overall N = 454	4 positive nodes N = 87	< 4 positive nodes N = 367
Age			
50	157 (35%)	28 (32%)	129 (35%)
>50	297 (65%)	59 (68%)	238 (65%)
Histology			
Ductal	363 (80%)	61 (70%)	302 (82%)
Lobular	91 (20%)	26 (30%)	65 (18%)
Tumor size			
2	237 (52%)	27 (31%)	210 (57%)
>2	217 (48%)	60 (69%)	157 (43%)
Tumor stage			
I	237 (52%)	27 (31%)	210 (57%)
II	178 (39%)	43 (49%)	135 (37%)
III	39 (9%)	17 (20%)	22 (6%)
LVI			
Present	202 (44%)	49 (56%)	153 (42%)
Absent	252 (56%)	38 (44%)	214 (58%)
Extranodal extension			
Present	203 (45%)	66 (76%)	137 (37%)
Absent	251 (55%)	21 (24%)	230 (63%)
Tumor grade			
I	107 (24%)	19 (22%)	88 (24%)
II	253 (56%)	49 (56%)	204 (56%)
III	94 (21%)	19 (22%)	75 (20%)
ER			
Positive	396 (87%)	76 (87%)	320 (87%)
Negative	58 (13%)	11 (13%)	47 (13%)
SLN removed			
1	172 (38%)	38 (44%)	134 (37%)
2	134 (30%)	27 (31%)	107 (29%)
3	75 (17%)	14 (16%)	61 (17%)
4	73 (16%)	8 (9%)	65 (18%)
Method of detection			
Frozen	369 (81%)	84 (97%)	285 (78%)
Serial H&E	54 (12%)	2 (2%)	52 (14%)
IHC	31 (7%)	1 (1%)	30 (8%)

	Overall N = 454	4 positive nodes N = 87	< 4 positive nodes N = 367
Positive SLN			
1	359 (79%)	50 (57%)	309 (84%)
2	78 (17%)	28 (32%)	50 (14%)
3	17 (4%)	9 (10%)	8 (2%)
Metastasis size			
< 2 mm	142 (31%)	8 (9%)	134 (37%)
2 mm	312 (69%)	79 (91%)	233 (63%)
Percentage of positive SLNs			
1–25	46 (10%)	1 (1%)	45 (12%)
26–50	166 (37%)	17 (20%)	149 (41%)
51–75	22 (5%)	6 (7%)	16 (4%)
76–100	220 (48%)	63 (72%)	157 (43%)

Discrimination statistics including the estimated AUC are summarized for each of the three models predicting the presence of 4 positive nodes versus < 4 positive nodes. For each model, the mean score was significantly different between the two groups, but the Rivers and Katz models demonstrated higher AUC values in our sample.

Table 2

	Calculated Score				
	N	AUC (95% CI)	Patients with 4 positive nodes Mean (SD) Median (Range)	Patients with < 4 positive nodes Mean (SD) Median (Range)	p-value
Rivers	454	0.81 (0.76, 0.86)	0.41 (0.27) 0.37 (0.01–0.91)	0.15 (0.18) 0.09 (0.004–0.85)	<0.0001
Chagpar	415	0.73 (0.67, 0.79)	4.0 (0.9) 4 (2, 6)	3.2 (0.8) 3 (1–6)	<0.0001
Katz	454	0.81 (0.76, 0.86)	–4.9 (17.1) –3 (–54–33)	–25.4 (16.7) –27 (–57–37)	<0.0001

Table 3

To assess model calibration, the observed proportion of patients in our validation sample with 4 nodes positive is reported with 95% confidence intervals across the range of possible scores and the corresponding probability estimates reported with each model. Observed proportions falling outside the model-based or reported range for the model are shown in bold.

N		Model-based or reported estimate of proportion with 4+ nodes positive	Observed proportion (n) with 4 nodes positive	95% CI for observed proportion
Rivers predicted probability				
<0.01	14	<0.01	0 (0)	(0, 0.23)
0.01–0.05	165	0.01–0.05	0.05 (9)	(0.03, 0.10)
0.06–0.10	38	0.06–0.10	0.08 (3)	(0.02, 0.21)
0.11–0.15	70	0.11–0.15	0.13 (9)	(0.06, 0.23)
0.16–0.20	16	0.16–0.20	0.25 (4)	(0.07, 0.52)
0.21–0.30	27	0.21–0.30	0.19 (5)	(0.06, 0.38)
0.31–0.40	50	0.31–0.40	0.38 (19)	(0.25, 0.53)
0.41–0.50	7	0.41–0.50	0.14 (1)	(0, 0.58)
0.51–0.75	51	0.51–0.75	0.47 (24)	(0.33, 0.62)
0.76–1.0	16	0.76–1.0	0.81 (13)	(0.54, 0.96)
Chagpar points				
1	2	0.01–0.02	0 (0)	(0, 0.84)
2	68	0.05–0.07	0.03 (2)	(0, 0.10)
3	167	0.15–0.16	0.10 (16)	(0.06, 0.15)
4	143	0.29–0.30	0.25 (36)	(0.18, 0.33)
5	30	0.49–0.52	0.40 (12)	(0.23, 0.59)
6	5	0.46–0.90	0.80 (4)	(0.28, 0.99)
Katz score				
–53	21	0.005	0.05 (1)	(0.001, 0.24)
–52 to –45	27	> 0.005 to 0.01	0.04 (1)	(0.001, 0.19)
–44 to –39	51	> 0.01 to 0.02	0.02 (1)	(0.001, 0.10)
–38 to –29	67	> 0.02 to 0.05	0.04 (3)	(0.01, 0.13)
–29	166	0.05 overall	0.04 (6)	(0.01, 0.08)
–28 to –22	61	> 0.05 to 0.10	0.10 (6)	(0.04, 0.20)
–21 to –14	86	> 0.10 to 0.20	0.20 (17)	(0.12, 0.30)
–28 to –14	147	> 0.05 to 0.20 overall	0.16 (23)	(0.10, 0.23)
–13 to –8	35	> 0.20 to 0.30	0.23 (8)	(0.10, 0.40)
–7 to –4	30	> 0.30 to 0.40	0.20 (6)	(0.08, 0.39)
–3 to 0	15	> 0.40 to 0.50	0.60 (9)	(0.32, 0.84)
–13 to 0	80	> 0.20 to 0.50 overall	0.29 (23)	(0.19, 0.40)
1 to 4	21	> 0.50 to 0.60	0.48 (10)	(0.26, 0.70)

	N	Model-based or reported estimate of proportion with 4+ nodes positive	Observed proportion (n) with 4 nodes positive	95% CI for observed proportion
5 to 8	13	> 0.60 to 0.70	0.54 (7)	(0.25, 0.81)
9 to 14	14	> 0.70 to 0.80	0.57 (8)	(0.29, 0.82)
15 to 22	7	> 0.80 to 0.90	0.86 (6)	(0.42, 1)
23 to 29	3	> 0.90 to 0.95	0.67 (2)	(0.09, 0.99)
30 to 39	3	> 0.95 to 0.98	0.67 (2)	(0.09, 0.99)
> 39	0	0.99	0	
> 0	61	> 0.50 overall	0.57 (35)	(0.44, 0.70)

Table 4

The sensitivity, specificity, and false negative percentages achieved in our sample are reported for each model at the clinically relevant cutpoints of 0.05 and 0.10 (Rivers and Katz models) and 1, 2, and 3 points for the Chagpar score.

Model	Predicted Probability / Points	Sensitivity	Specificity	False negatives (%)
Rivers	0.05 vs > 0.05	78/87 (90%)	170/367 (46%)	9 (10%)
	0.10 vs > 0.10	75/87 (86%)	205/367 (56%)	12 (14%)
Chagpar	1 point vs > 1 point	70/70 (100%)	2/345 (1%)	0 (0%)
	2 points vs > 2 points	68/70 (97%)	68/345 (20%)	2 (3%)
	3 points vs > 3 points	52/70 (74%)	219/345 (63%)	18 (26%)
Katz	0.05 (score -29) vs > 0.05	81/87 (93%)	160/367 (44%)	6 (7%)
	0.10 (score -22) vs > 0.10	75/87 (86%)	215/367 (59%)	12 (14%)