Breast Cancer Res Treat (2012) 134:661–670
DOI 10.1007/s10549-012-2109-2

PRECLINICAL STUDY

Predictions of the pathological response to neoadjuvant chemotherapy in patients with primary breast cancer using a data mining technique

M. Takada · M. Sugimoto · S. Ohno · K. Kuroi · N. Sato · H. Bando · N. Masuda · H. Iwata · M. Kondo · H. Sasano · L. W. C. Chow · T. Inamoto · Y. Naito · M. Tomita · M. Toi

Received: 21 March 2012 / Accepted: 22 May 2012 / Published online: 12 June 2012

Abstract Nomogram, a standard technique that utilizes multiple characteristics to predict efficacy of treatment and likelihood of a specific status of an individual patient, has been used for prediction of response to neoadjuvant chemotherapy (NAC) in breast cancer patients. The aim of this study was to develop a novel computational technique to predict the pathological complete response (pCR) to NAC in primary breast cancer patients. A mathematical model using alternating decision trees, an epigone of decision tree, was developed using 28 clinicopathological variables that were retrospectively collected from patients treated with NAC ($n = 150$), and validated using an independent dataset from a randomized controlled trial ($n = 173$). The model selected 15 variables to predict the pCR with yielding area under the receiver operating characteristics curve (AUC) values of 0.766 [95 % confidence interval (CI)], 0.671–0.861, P value < 0.0001) in cross-validation using training dataset and 0.787 (95 % CI 0.716–0.858, P value < 0.0001) in the validation dataset. Among three subtypes of breast cancer, the luminal subgroup showed the best discrimination (AUC = 0.779, 95 % CI 0.641–0.917, P value = 0.0059). The developed model (AUC = 0.805, 95 % CI 0.716–0.894, P value < 0.0001) outperformed multivariate logistic regression (AUC = 0.754, 95 % CI 0.651–0.858, P value = 0.00019) of validation datasets without missing values ($n = 127$). Several analyses, e.g.

Electronic supplementary material The online version of this article (doi:10.1007/s10549-012-2109-2) contains supplementary material, which is available to authorized users.

M. Takada · M. Toi (✉)
Department of Breast Surgery, Graduate School of Medicine, Kyoto University, 54 Kawaracho, Shogoin, Sakyo-ku, Kyoto 606-8507, Japan
e-mail: toi@kuhp.kyoto-u.ac.jp

M. Takada
Japan Society for the Promotion of Science, Tokyo, Japan

M. Sugimoto
Medical Innovation Centre, Kyoto University Graduate School of Medicine, Kyoto University, Kyoto, Japan

M. Sugimoto · Y. Naito · M. Tomita
Institute for Advanced Biosciences, Keio University, Tsuruoka, Yamagata, Japan

S. Ohno
Department of Breast Oncology, National Kyushu Cancer Centre, Fukuoka, Japan

K. Kuroi
Department of Surgery, Tokyo Metropolitan Cancer and Infectious Diseases Centre, Komagome Hospital, Tokyo, Japan

N. Sato
Department of Surgery, Niigata Cancer Centre Hospital, Niigata, Japan

H. Bando
Department of Breast and Endocrine Surgery, Faculty of Medicine, University of Tsukuba, Tsukuba, Ibaraki, Japan

N. Masuda
Department of Surgery, Breast Oncology, Osaka National Hospital, Osaka, Japan

H. Iwata
Department of Breast Oncology, Aichi Cancer Centre, Nagoya, Aichi, Japan

M. Kondo
Department of Health Care Policy and Management, Graduate School of Comprehensive Human Sciences, University of Tsukuba, Tsukuba, Ibaraki, Japan

H. Sasano
Department of Pathology, Tohoku University Hospital and School of Medicine, Sendai, Miyagi, Japan

bootstrap analysis, revealed that the developed model was insensitive to missing values and also tolerant to distribution bias among the datasets. Our model based on clinicopathological variables showed high predictive ability for pCR. This model might improve the prediction of the response to NAC in primary breast cancer patients.

Keywords Breast cancer · Data mining · Neoadjuvant chemotherapy · Nomogram · Prediction model

Introduction

Neoadjuvant chemotherapy (NAC) is the administration of chemotherapy before surgical treatment of cancer. The clinical advantages of NAC include tumour size reduction, which improves the breast conservation rate, and determination of chemosensitivity to help design later adjuvant therapy [1]. Several meta-analyses have revealed that patients with pathological complete response (pCR) after NAC showed higher survival rates than those without pCR [1–4], indicating that pCR might represent a surrogate prognostic indicator in these patients. Thus, predicting pCR using information collected before NAC has been proposed, with the most commonly used predictive factors including oestrogen receptor (ER) status, human epidermal growth factor receptor 2 (HER2/neu) status, histological grade and proliferative activity [5–7]. Recent studies showed that the sensitivity to chemotherapy differs according to cancer phenotype classified mainly by ER and HER2 status [8–10]. Luminal A subtype (ER-positive, HER2-negative and low-grade or low-proliferative phenotype) exhibited lower sensitivity to chemotherapy despite better prognosis than other phenotypes, and hormonal therapy alone is the preferred treatment for this subtype [11].

Nomograms, which integrate clinical and pathological variables using multiple logistic regression (MLR), have been developed and are well validated to predict pCR after NAC [12, 13]. These tools are now widely used by oncologists via sophisticated web interfaces. In comparative studies using prospective cohorts, nomograms showed similar performance to molecular tests designed to predict pCR following NAC [14, 15]. The gene signatures to predict prognosis were also expected to predict chemosensitivity [16–19], although the predictive ability was limited in those studies. Thus, new prediction tools, even with the use of molecular/clinicopathological factors, are now needed.

MLR has several limitations. First, it can deal with only few independent variables to avoid over-fitting to the given datasets. Second, MLR is sensitive to missing values, a frequent occurrence in retrospectively collected data. Third, MLR cannot tolerate the distribution bias of variables among multiple datasets (usually obtained from different institutes), which reduces its generalizability. Thus, we used a data mining technique to address the following problems: (1) limits in the number of variables that can be included in a model, (2) missing values and (3) bias among datasets. Using alternating decision tree (ADTree), an accurate and versatile decision-tree type data mining method [20], we developed and validated a mathematical model to predict pCR after NAC in patients with primary breast cancer.

Materials and methods

Participants and treatments

The study protocol was approved by the institutional review board of Kyoto University Hospital. We disclosed the details of the study to all of the participants in lieu of obtaining informed consent because the Japanese ethics guidelines for epidemiologic research allow observational studies to use anonymous clinical data after disclosing the study details to the potential participants.

We included patients who had participated in the Organisation for Oncology and Translational Research (OOTR) N003 trial. This was a randomized trial of patients with operable breast cancer treated with docetaxel with or without capecitabine after four cycles of NAC consisting of 5-fluorouracil, epirubicin and cyclophosphamide (FEC) (UMIN ID: C000000322, http://www.umin.ac.jp/ctr/index.htm). Patients who received the same chemotherapy regimen in regular clinical practice were also included in this study. Only patients with a tumour size of $\leq$5 cm and who had completed $\geq$75 % of the planned courses of NAC were included.

First, we conducted an exploratory analysis using a dataset of 58 patients collected consecutively from Tokyo Metropolitan Cancer and Infectious Diseases Centre

L. W. C. Chow
Comprehensive Centre for Breast Diseases, UNIMED Medical Institute, Hong Kong, China

T. Inamoto
Department of Breast Surgery, Tenri Hospital, Tenri, Nara, Japan

Y. Naito · M. Tomita
Faculty of Environment and Information Studies, Keio University, Fujisawa, Kanagawa, Japan

Y. Naito · M. Tomita
Graduate School of Media and Governance, Keio University, Fujisawa, Kanagawa, Japan

Komagome Hospital. Additional patients were then collected from Osaka National Hospital and Tsukuba University Hospital to develop the prediction model. The training dataset consisted of 150 patients treated at the three institutions from 2005 to 2009. This included 89 patients who participated in the OOTR N003 trial. The newly developed prediction model was applied to an external validation dataset consisting of 173 patients from the OOTR N003 trial. This validation dataset was obtained from three randomly selected institutions that had participated in the OOTR N003 trial (Niigata Cancer Centre Hospital, National Kyushu Cancer Centre and Aichi Cancer Centre).

All of the patients included in this study received the same treatment protocol, consisting of four courses of FEC (5-fluorouracil 500 mg/m^2, epirubicin 100 mg/m^2 and cyclophosphamide 500 mg/m^2, i.v., every 3 weeks) followed by four courses of docetaxel (75 mg/m^2, i.v., every 3 weeks) with or without capecitabine (1,650 mg/m^2/day, oral administration, for 14 days every 3 weeks).

Data collection

Data for 28 clinicopathological variables were retrospectively collected from databases maintained at each institution (Table 1). All of the mammography and ultrasonography images were reviewed by physicians certified in imaging diagnosis by the relevant accreditation organizations in Japan. The Response Evaluation Criteria in Solid Tumours (RECIST) were used to classify the clinical response to NAC. According to the protocol of the OOTR N003 trial, the clinical response was evaluated after both the FEC treatment (i.e. the clinical response after anthracycline) and the taxane-containing regimen (i.e. the clinical response after taxane). Information pertaining to histological type, ER status, progesterone receptor (PgR) status, HER2 status and histological/nuclear grade of needle biopsy specimens were collected from the original pathology report of each patient. At each institution, the surgical specimen obtained following NAC was serially sectioned, stained with haematoxylin and eosin (H&E) and diagnosed by experienced pathologists. pCR was defined as the absence of residual invasive cancer cells in the breast and the axillary lymph nodes (ypT0/is + ypN0).

Data analysis

For statistical analyses, we quantitatively graded the variables using the criteria established by a committee of specialists from the fields of breast surgery, diagnostic radiology and pathology.

The ADTree model was developed using the training dataset and validated using the independent validation

Table 1 Characteristics of patients in the training and validation datasets

Variables	Training dataset		Validation dataset		*P* value[d]
	n	%	*n*	%	
No. of patients	150		173		
Age (years)					
Median	50		48		0.131
Range	(27–71)		(29–68)		
Body mass index (kg/m^2)					
Median	22.7		21.2		0.0004
Range	(16.9–35.8)		(15.6–43)		
Menopausal status					
Pre-menopause	75	50.0	105	60.7	0.054
Post-menopause	75	50.0	68	39.3	
Physical examination					
Palpable lump					
Yes	149	99.3	166	96.0	0.136
No	1	0.7	5	2.9	
Unknown	0	0	2	1.2	
Skin dimpling					
Yes	31	20.7	28	16.2	0.281
No	116	77.3	143	82.7	
Unknown	3	2.0	2	1.2	
Mammography					
Presence of a mass					
Yes	92	61.3	113	65.3	0.102
Focal asymmetry	19	12.7	35	20.2	
No	26	17.3	20	11.6	
Unknown	13	8.7	5	2.9	
Presence of calcifications					
Yes	45	30.0	80	46.2	0.035
No	87	58.0	88	50.9	
Unknown	15	10.0	5	2.9	
Architectural distortion					
Yes	27	18.0	22	12.7	0.110
No	108	72.0	145	83.8	
Unknown	15	10.0	6	3.5	
Ultrasonography					
Presence of masses					
Yes	148	98.7	172	99.4	0.480
No	2	1.3	1	0.6	
Maximum tumour size (mm)					
Median	26		29		0.012
Range	(11–48)		(11–49)		
Unknown	0	0	8	4.7	
Depth/width ratio					
Median	0.67		0.58		<0.0001
Range	(0.23–2.06)		(0.22–1.18)		

Table 1 continued

Variables	Training dataset		Validation dataset		*P* value[d]
	n	%	*n*	%	
Unknown	3	2.0	21	12.2	
Echogenic halo					
Yes	68	45.9	56	32.6	0.026
No	79	53.4	109	63.4	
Unknown	1	0.7	7	4.1	
Posterior acoustic features					
Enhancement	33	22.3	61	35.5	0.024
None	75	50.7	66	38.4	
Shadowing	37	25.0	41	23.8	
Unknown	3	2.0	4	2.3	
Interruption of the anterior border of the mammary gland					
Yes	123	83.1	146	84.9	0.694
No	22	14.9	23	13.4	
Unknown	3	2.0	3	1.7	
Histological type					
Invasive ductal carcinoma	146	97.3	170	98.3	0.566
Invasive lobular carcinoma	4	2.7	3	1.7	
ER status[a]					
Positive	105	70.0	102	59	0.024
Negative	42	28.0	70	40.5	
Unknown	3	2.0	1	0.6	
PgR status[a]					
Positive	69	46.0	84	48.6	0.735
Negative	78	52.0	88	50.9	
Unknown	3	2.0	1	0.6	
HER2 status[b]					
Positive	19	12.7	38	22.0	0.026
Negative	125	83.3	127	73.4	
Unknown	6	4.0	8	4.6	
Triple-negative phenotype[c]					
Yes	31	20.7	44	25.4	0.293
No	113	75.3	121	70.0	
Unknown	6	4.0	8	4.6	
Histological/nuclear grade					
1/2	94	62.7	89	51.4	0.446
3	49	32.7	38	22.0	
Unknown	7	4.7	46	26.6	
Mitotic index					
1	68	45.3	68	39.3	0.137
2	40	26.7	23	13.3	
3	33	22.0	21	12.1	
Unknown	9	6.0	61	35.3	
Treatment regimen					
FEC-T	108	72.0	94	54.3	0.001
FEC-TX	42	28.0	79	45.7	

NS not significant, *NC* not collected, *FEC* 5-fluorouracil + epirubicin + cyclophosphamide, *T* taxane, *TX* taxane + capecitabine

[a] ER-positive or PgR-positive was defined as 10 % or more of cells with positive staining or Allred score of 3 or more

[b] HER2-positive was defined as a score of 3+ on immunohistochemical testing or a positive score on fluorescence in situ hybridization testing

[c] Triple negative was defined as ER, PgR and HER2 negative

[d] χ^2 test, Mann–Whitney *U* or *t* test

dataset. To enhance model accuracy, we used ensemble methods: multiple ADTree models were developed and the mean prediction of these models was used as the final prediction [21]. The model was optimized by cross-validation (CV) and the area under the receiver operating characteristics curve (AUC) for discriminating pCR from non-pCR was determined.

The importance of variables in the ADTree model was evaluated based on the decrease in prediction accuracy (AUC values) by replacing the actual value with a random value for each variable (sensitivity analysis). To evaluate the significance of missing values in the developed model, the missing values were replaced with random values and the decrease in AUC value was assessed (missing value analysis). The prediction accuracy was evaluated using a smaller number of ADTrees in the developed model than the optimized number (pruning analysis). To elucidate the relationship between generalizability and variable distribution bias between the training and the validation dataset, we integrated all of the data and randomly split it into two datasets. The ADTree model was developed using one of these datasets and validated using the other dataset (random split analysis). Each analysis was repeated 200 times with different random values.

We also developed an MLR model using the training dataset. Details of this model and the software used are described in the Supplementary Materials and methods.

Results

The clinicopathological variables of each dataset are summarized in Table 1. The training dataset included more ER-positive or HER2-negative patients compared with the validation dataset (*P* value = 0.024 and 0.026, respectively). Overall, 16 % of patients in the training dataset and 22.5 % in the validation dataset achieved pCR

Table 2 Treatment outcomes of the training and validation datasets

Outcomes	Training dataset		Validation dataset		*p* value[a]
	n	%	*n*	%	
No. of patients	150		173		
Clinical response after anthracycline treatment					
CR + PR	99	66.0	151	87.3	<0.0001
SD	50	33.3	21	12.1	
PD	0	0	1	0.6	
Unknown	1	0.7	0	0	
Clinical response after taxane treatment					
CR + PR	128	85.3	164	94.8	0.006
SD	21	14	7	4	
PD	1	0.7	2	1.2	
Breast surgery					
Mastectomy	32	21.3	53	30.6	0.058
Breast-conserving surgery	118	78.7	120	69.4	
pCR (ypT0/is + ypN0)					
Yes	24	16.0	39	22.5	0.139
No	126	84.0	134	77.5	

CR complete response, *PR* partial response, *SD* stable disease, *PD* progression disease

[a] χ^2 test

(*P* value = 0.139) (Table 2). The rates of pCR and breast conservation were not significantly different between the institutions (*P* value = 0.06 and 0.30, respectively). The clinical responses after anthracycline and taxane, however, were significantly lower in the training dataset than in the validation dataset (*P* value < 0.0001 and *P* = 0.006, respectively).

The selected model showing the best AUC value in the CVs contained 19 ADTrees with three variables on each tree (Fig. 1a; Supplementary Fig. S1). In total, 15 variables were included: three general [body mass index (BMI), menopausal status and the presence of skin dimpling], five ultrasonographic (maximum tumour size, tumour depth/width ratio, echogenic halo, interruption of the anterior border of the mammary gland and posterior acoustic features), three mammographic (the presence of calcifications, the presence of a mass and architectural distortion), and four pathological variables (mitotic index and the status of ER, PgR and HER2). The method used to calculate the probability of pCR using this model is shown in Fig. 1b and Supplementary Fig. S2.

The receiver operating characteristics (ROC) curves and the dot-plots of the pCR for each dataset are shown in Fig. 2. The AUC values were 0.766 (95 % CI 0.671–0.861, *P* value < 0.0001) in the CV using the training dataset and 0.787 (95 % CI 0.716–0.858, *P* value < 0.0001) using the validation dataset. The model could discriminate pCR from non-pCR patients at significant levels in both the training and the validation datasets (*P* value < 0.0001). When the threshold for a low risk of pCR was defined as 20 % for example, the false-negative rate was 7.7 % and the negative predictive value was 95.9 % using the validation dataset. The AUC of bootstrap analysis (200 repetitions) performed to obtain unbiased estimates was 0.791 (95 % CI: 0.786–0.796) using the validation dataset.

To assess the prediction ability by integrating early clinical response, we developed a MLR model comprising two variables; the predicted probability of pCR determined by ADTree and the clinical responses after anthracycline or taxane treatment. The accuracy of the ADTree model was enhanced by including the clinical response to NAC (Supplementary Fig. S3). The AUC values for the validation datasets were 0.820 (95 % CI 0.757–0.883, *P* value < 0.0001) and 0.855 (95 % CI 0.794–0.916, *P* value < 0.0001) after including the clinical responses after anthracycline and taxane treatment, respectively.

We evaluated the discriminative ability of our model in three subgroups of patients with luminal (ER-positive and any HER2 status; *n* = 102), HER2-positive (ER-negative and HER2-positive; *n* = 24) or triple-negative (ER- and HER2-negative; *n* = 44) patterns of receptor expression. The model showed significant discrimination of the luminal subgroup (*P* value = 0.0059), poor discrimination of the triple-negative subgroup (*P* value = 0.743) and moderate discrimination of the HER2-positive subgroup (*P* value = 0.074) (Fig. 3).

The sensitivity analysis revealed large decreases in AUC values when ER, PgR, HER2 and echogenic halo were randomly replaced, indicating high importance of these factors. On the other hand, the AUC values were hardly changed when imaging findings were randomly replaced, indicating little importance of these parameters (Fig. 4a). In missing value analysis, the median AUC was 0.786 (95 % CI 0.785–0.787) using validation datasets including patients with at least one missing value (*n* = 121). As another approach to evaluate the importance of the variables included in the model, pruning analysis was performed to reduce the number of trees, which also reduces the number of variables. In this analysis, the AUC value for the overall dataset remained high (>0.78) when the number of trees was >4 and the number of variables was >6 (Fig. 4b). When the number of variables was reduced to 8 for example, discrimination of ER-positive and, particularly, HER2-positive subtypes deteriorated (*P* value = 0.0266 and *P* value = 0.725), respectively (Supplementary Fig. S4). In random split analysis, the median AUC value was 0.776 (95 % CI 0.773–0.776), almost identical to the AUC value obtained using the original validation dataset (0.787).

To compare ADTree and MLR, a MLR model was developed using our training dataset and consisted of ER,

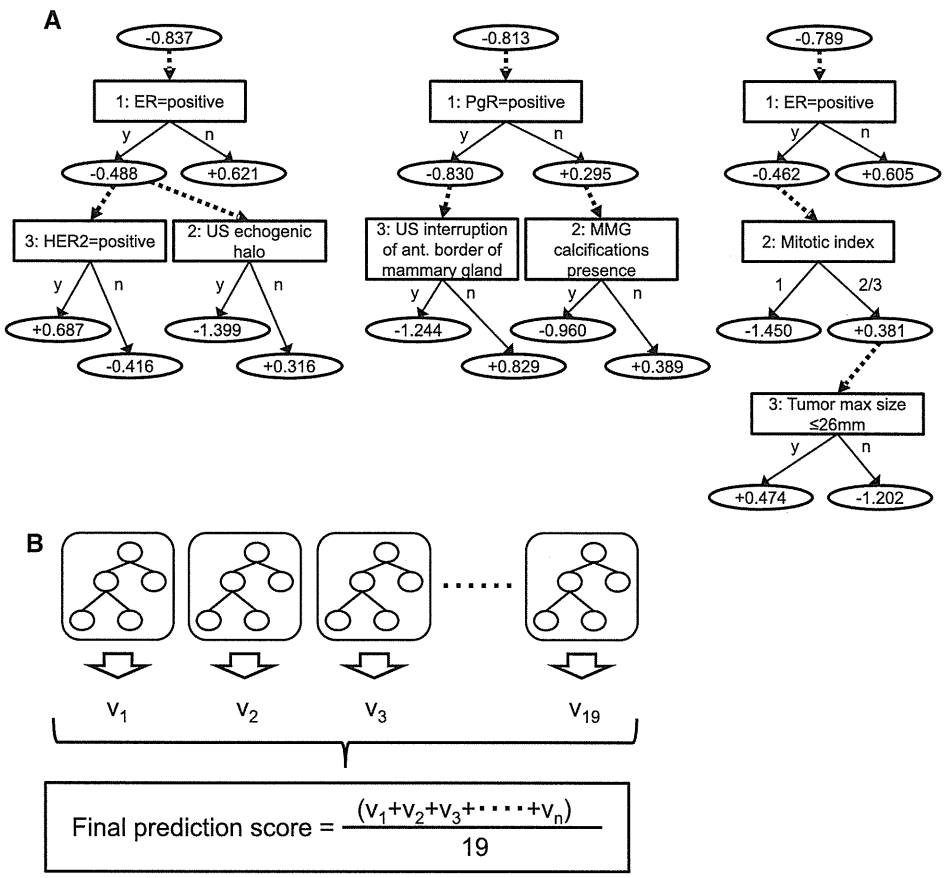

Fig. 1 ADTree-based prediction model. **a** Part of the developed ADTree models. The final prediction model consisted of 19 ADTree-based prediction models; the other 16 models are depicted in Supplementary Fig. S1. The method used to calculate the prediction score in each model is described in Supplementary Fig. S2. **b** The final prediction score was calculated by calculating the mean score of the 19 ADTree-based models. V_1, V_2 and V_{19} indicate the prediction scores of each ADTree. The probability of pCR (%) was determined using the formula $(score_{pred} - score_{min})/(score_{max} - score_{min}) \times 100$, where $score_{pred}$, $score_{min}$ and $score_{max}$ are the predicted and theoretical minimum and maximum final scores, respectively

histological/nuclear grade and interruption of the anterior border of the mammary gland on US. This MLR model yielded an AUC value of 0.754 (95 % CI 0.651–0.858, P value = 0.00019) using a subset of the validation dataset (n = 127; 46 cases were excluded because of missing values). The ADTree model outperformed the MLR model by yielding an AUC of 0.805 (95 % CI 0.716–0.894, P value < 0.0001) using the same patient dataset.

Discussion

Here, we developed a prediction model for pCR after NAC using ADTree and analyzed the model using several analyses. The validation dataset was from the OOTR N003 trial, in which patients received FEC treatment and were randomly assigned to four cycles of docetaxel alone (FEC-T) or four cycles of docetaxel plus capecitabine (FEC-TX). In the validation dataset, the AUC values for the FEC-T and FEC-TX groups were 0.789 (P value < 0.0001) and 0.788 (P value = 0.0003), respectively. The GeparQuattro study reported that adding capecitabine to preoperative docetaxel after four courses of epirubicin and cyclophosphamide did not improve the rate of pCR [22]. Accordingly, our model showed a similar performance for both the regimens.

The proportion of patients with the luminal, HER2-negative subtype was higher in the training dataset than in the validation dataset (65 and 50 %, respectively; P value = 0.008), which indicated that the training dataset included more patients with potentially chemo-insensitive cancers compared with the validation dataset. This disproportion may have led to the difference in the clinical response rate between the training and validation datasets (66 vs. 87 % after anthracycline and 85 vs. 95 % after taxane, respectively; Table 2). However, the training

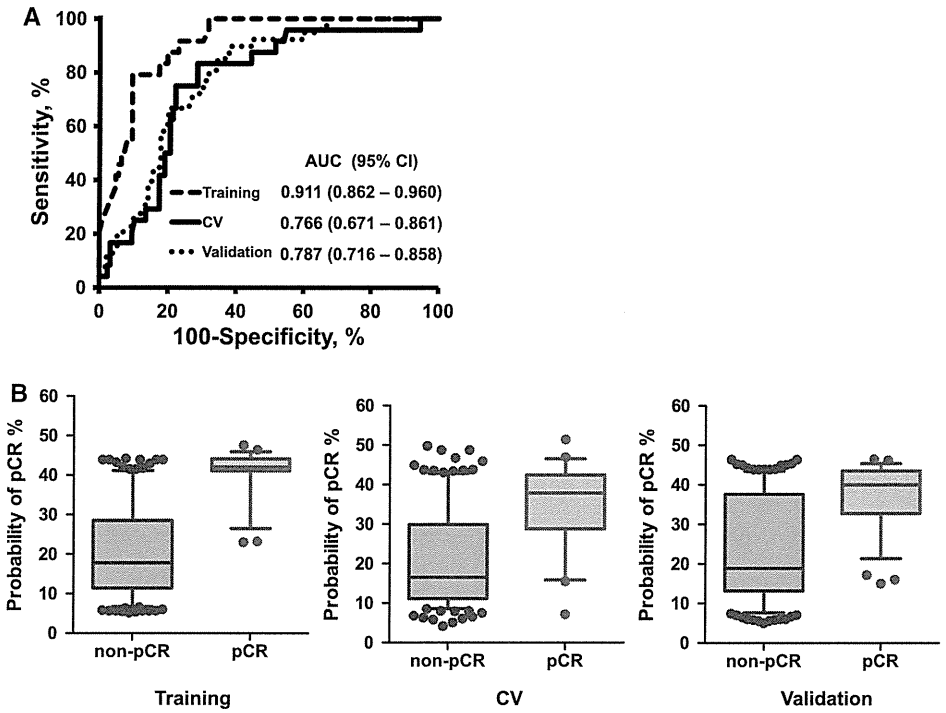

Fig. 2 **a** ROC curves and **b** the distribution of the predicted probabilities. **a** ROC curves of the prediction model. The area under the ROC curve (AUC) values were 0.766 (95 % CI 0.671–0.861, P value < 0.0001) in the CV using the training dataset, and 0.787 (95 % CI 0.716–0.858, P value < 0.0001) using the validation dataset. **b** Box plots showing the distribution of the predicted probabilities of pCR determined using our model. The *box* and *whiskers* represent the 10th, 25th, 50th, 75th and 90th percentiles, and the data below the 10th and above the 90th percentile were plotted. In each figure, the *boxes* on the *left side* of the graph represent the patients without pCR (non-pCR), and the *boxes* on the *right side* represent the patients who did achieve pCR. The model was able to discriminate pCR from non-pCR patients at significant levels (P value < 0.0001)

dataset included patients with smaller tumours compared with the validation dataset (median diameter: 26 and 29 mm, respectively; P value = 0.012; Table 1). Therefore, the differences in the clinical response rate may not be associated with differences in the pathological response rate or breast-conserving rate. Although the unequal distribution of cancer subtypes between the two datasets may affect the generalizability of our model, the result of random split analysis showed that this discrepancy hardly affected the predictive performance of our model.

There are several criteria used to define pCR after NAC. Here, we defined pCR as the absence of residual invasive cancer cells in the breast or the lymph nodes (ypT0/is + ypN0). pCR has also been defined as the complete disappearance of cancer cells from the breast and the lymph nodes (ypT0 + ypN0), and as the absence of residual cancer cells in the breast tissue, regardless of lymph node status (ypT0 + ypNX) [23]. Therefore, we evaluated the predictive performance of ADTree model using these three definitions. The AUC values were 0.728 (95 % CI 0.640–0.817; P value = 0.0002) for ypT0 + ypN0 and 0.786 (95 % CI 0.705–0.867; P value < 0.0001) for ypT0 + ypNX in the validation dataset. Although our model identified patients with and without pCR at significant levels using both the definitions, the accuracy of our model was decreased when the ypT0 + ypN0 definition was used. The rate of pCR determined by the ypT0 + ypN0 definition was lower than that for ypT0/is + ypN0 (10 vs. 16 % for the training dataset and 15 vs. 22.5 % for the validation dataset). Therefore, further evaluation using a larger dataset is needed.

It has been reported that an early clinical response to NAC might be predictive of pCR [2, 24]. As expected, the AUC values obtained using validation datasets increased to 0.820 after including the clinical responses after anthracycline treatment. Therefore, the ADTree model can provide highly accurate prediction for pCR by integrating the early clinical response to NAC.

The sensitivity analysis suggested that ER, PgR and HER2 were more important than the other variables. Among the imaging features that were generally less important than these three features, echogenic halo was the most important. Recent studies have indicated that the subgroup, mainly classified using ER and HER2 status, is strongly associated with the pathological response to NAC

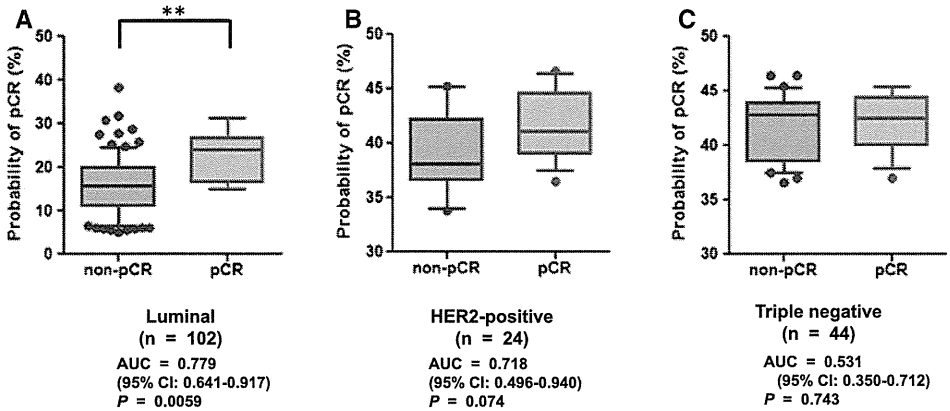

Fig. 3 Box plots showing the distribution of the predicted probabilities of pCR using the prediction model in each subgroup. In each figure, the *boxes* on the *left side* of the graph represent the patients without pCR (non-pCR), and the *boxes* on the *right side* represent the patients who did achieve pCR. Each subgroup was defined as follows: **a** luminal type (ER-positive and any HER2 status; $n = 102$); **b** HER2-positive type (ER-negative and HER2-positive; $n = 24$) and **c** triple-negative type (ER- and HER2-negative; $n = 44$)

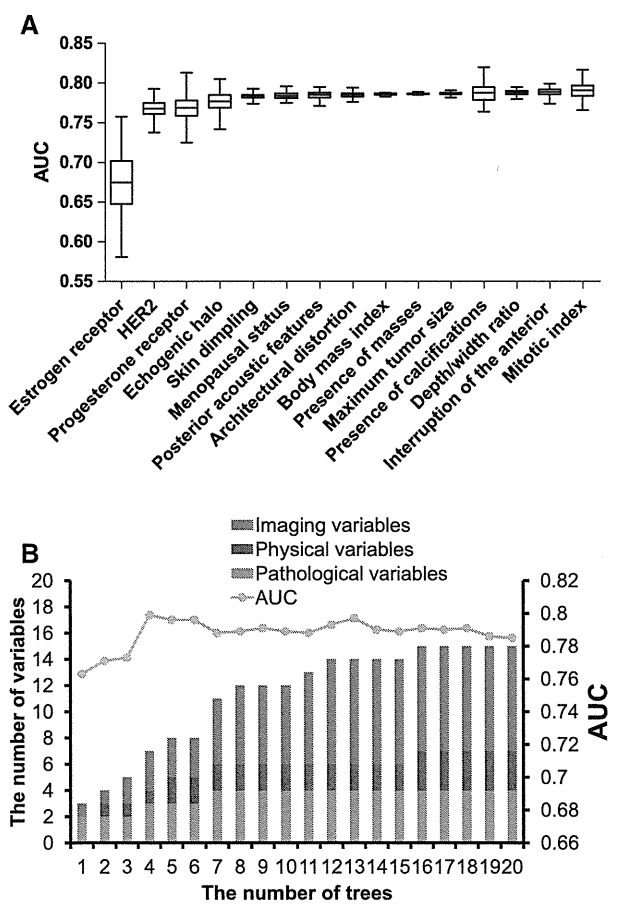

Fig. 4 **a** Sensitivity analysis and **b** pruning analysis using the validation dataset. **a** Box plots show the maximum and minimum, the 25th and 75th percentiles, and the median AUC values. **b** The horizontal axis shows the number of trees (bagging number). For each number of trees, a *bar graph* shows the number of variables and a *line graph* shows the AUC values

[8–10, 18]; however, predicting pCR within each phenotype is still challenging. The model showed relatively higher AUC values in luminal, HER2-negative and HER2-positive subtypes (Fig. 3). Meanwhile, the model did not function for the triple-negative phenotype (AUC = 0.531), which means the clinicopathological variables collected in this study show limited potential to predict chemosensitivity in this phenotype. One possible reason of this low AUC is the heterogeneity of triple-negative populations [25]. Thus, the identification of new variables, including novel genomic and epigenetic markers, and new models integrating these variables are needed to overcome this limitation.

The missing value analysis yielded an AUC value of 0.786 (95 % CI 0.785–0.787). This value was not much different from the result of bootstrap analysis (AUC = 0.791). It is particularly noteworthy that the difference between the upper and lower 95 % CI values was very small, indicating low sensitivity of the developed model to missing values. This is one of the beneficial features of the ensemble technique used in this study.

In the pruning analysis, the AUC values for the validation dataset improved according to the number of ADTrees in the prediction model (Fig. 4b). Reducing the number of trees to six eliminated many variables corresponding to imaging findings. Although the AUC value of the whole validation dataset remained high, the predictive performance for the luminal and ER-negative/HER2-positive subtypes decreased (Supplementary Fig. S4), which indicates that the variables derived from imaging findings might contribute the most to chemosensitivity prediction in both the subgroups.

Nomograms that use MLR to predict response to NAC have already been introduced [12, 13, 26]. In our study,

ADTree outperformed MLR using an identical dataset. MLR offers some advantages, particularly the use of fewer variables, which facilitates data collection and interpretation of the model. These features of each modelling method represent trade-offs that should be considered when applying the models. The combined use of multiple prediction models could enhance predictive accuracy [27]. We are currently testing the combination of our model and available nomograms in a prospective study.

There are several limitations of this study. Validation using larger databases will more accurately assess the model. The use of many features obtained from imaging studies or physical examination would reduce the number of users depending on the availability of the features. The datasets obtained from multiple institutes would contribute to strict evaluation of the model's versatility whereas such datasets sometimes introduce institute-dependent bias. In this study, we used information from individual pathology reports and the central pathology review is more preferable to evaluate the features in a single criteria. A Web-based interface to facilitate data input and prediction analysis, like the MD Anderson Cancer Centre nomogram, and an automated system to update the model will also be useful. Biomarkers of tumour response, particularly those obtained from midcourse biopsy samples, may increase the predictive accuracy. Integration with subtype-specific biomarkers is also needed to improve the accuracy of the developed model.

In conclusion, we have established a new ADTree-based method to predict pCR after NAC using variables readily collected before NAC. The model could use larger number of variables with keeping high generalization ability and showed the outperformed prediction accuracy compared with MLR as well as was tolerant to missing values and distribution bias in the datasets.

Acknowledgments We thank the doctors and data managers for data collection. We also thank the patients who participated in this study. This study was funded by research grants from the Ministry of Health, Labour and Welfare ("A study on the construction of an algorithm for multimodal therapy with biomarkers for primary breast cancer by formulation of a decision-making process", led by MT, No. H18-3JIGAN-IPPAN-007 and "Reduction and lowering of recurrence risk, toxicity and pharmacoeconomic cost by prediction of efficacy for anticancer agents in breast cancer patients", led by MT; No. H22-GANRINSHO-IPPAN-039), research funds from the Yamagata Prefectural Government and Tsuruoka City, and an International Internship Grant from the Global COE project "Centre for Frontier Medicine", Kyoto University. This study was also supported by the program "Raising Proficient Oncologists" administered by the Japanese Ministry of Education, Culture, Sports, Science and Technology.

Disclosure Dr. Hiroji Iwata has received honoraria from Chugai Pharmaceutical Co., Ltd, Japan. All remaining authors have declared no conflicts of interest.

References

1. van der Hage JH, van de Velde CCJH, Mieog SJSD (2007) Preoperative chemotherapy for women with operable breast cancer. Cochrane Database Syst Rev CD005002. doi:10.1002/14651858.CD005002.pub2
2. Bear HD, Anderson S, Smith RE, Geyer CE Jr, Mamounas EP, Fisher B, Brown AM, Robidoux A, Margolese R, Kahlenberg MS, Paik S, Soran A, Wickerham DL, Wolmark N (2006) Sequential preoperative or postoperative docetaxel added to preoperative doxorubicin plus cyclophosphamide for operable breast cancer: National Surgical Adjuvant Breast and Bowel Project Protocol B-27. J Clin Oncol 24:2019–2027. doi:10.1200/JCO.2005.04.1665
3. Fisher B, Bryant J, Wolmark N, Mamounas E, Brown A, Fisher ER, Wickerham DL, Begovic M, DeCillis A, Robidoux A, Margolese RG, Cruz AB Jr, Hoehn JL, Lees AW, Dimitrov NV, Bear HD (1998) Effect of preoperative chemotherapy on the outcome of women with operable breast cancer. J Clin Oncol 16:2672–2685
4. Mauri D, Pavlidis N, Ioannidis JP (2005) Neoadjuvant versus adjuvant systemic treatment in breast cancer: a meta-analysis. J Natl Cancer Inst 97:188–194. doi:10.1093/jnci/dji021
5. Chang J, Powles TJ, Allred DC, Ashley SE, Clark GM, Makris A, Assersohn L, Gregory RK, Osborne CK, Dowsett M (1999) Biologic markers as predictors of clinical outcome from systemic therapy for primary operable breast cancer. J Clin Oncol 17:3058–3063
6. Colleoni M, Viale G, Goldhirsch A (2009) Lessons on responsiveness to adjuvant systemic therapies learned from the neoadjuvant setting. Breast 18(Suppl 3):S137–S140. doi:10.1016/S0960-9776(09)70289-9
7. von Minckwitz G, Untch M, Nuesch E, Loibl S, Kaufmann M, Kummel S, Fasching PA, Eiermann W, Blohmer JU, Costa SD, Mehta K, Hilfrich J, Jackisch C, Gerber B, du Bois A, Huober J, Hanusch C, Konecny G, Fett W, Stickeler E, Harbeck N, Muller V, Juni P (2011) Impact of treatment characteristics on response of different breast cancer phenotypes: pooled analysis of the German neo-adjuvant chemotherapy trials. Breast Cancer Res Treat 125:145–156. doi:10.1007/s10549-010-1228-x
8. Colleoni M, Viale G, Zahrieh D, Bottiglieri L, Gelber RD, Veronesi P, Balduzzi A, Torrisi R, Luini A, Intra M, Dellapasqua S, Cardillo A, Ghisini R, Peruzzotti G, Goldhirsch A (2008) Expression of ER, PgR, HER1, HER2, and response: a study of preoperative chemotherapy. Ann Oncol 19:465–472. doi:10.1093/annonc/mdm509
9. Darb-Esfahani S, Loibl S, Muller BM, Roller M, Denkert C, Komor M, Schluns K, Blohmer JU, Budczies J, Gerber B, Noske A, du Bois A, Weichert W, Jackisch C, Dietel M, Richter K, Kaufmann M, von Minckwitz G (2009) Identification of biology-based breast cancer types with distinct predictive and prognostic features: role of steroid hormone and HER2 receptor expression in patients treated with neoadjuvant anthracycline/taxane-based chemotherapy. Breast Cancer Res 11:R69. doi:10.1186/bcr2363
10. Jones RL, Salter J, A'Hern R, Nerurkar A, Parton M, Reis-Filho JS, Smith IE, Dowsett M (2010) Relationship between oestrogen receptor status and proliferation in predicting response and long-term outcome to neoadjuvant chemotherapy for breast cancer. Breast Cancer Res Treat 119:315–323. doi:10.1007/s10549-009-0329-x
11. Goldhirsch A, Wood WC, Coates AS, Gelber RD, Thürlimann B, Senn HJ (2011) Strategies for subtypes—dealing with the diversity of breast cancer: highlights of the St. Gallen International Expert Consensus on the Primary Therapy of Early Breast

Cancer 2011. Ann Oncol 22:1736–1747. doi:10.1093/annonc/mdr304
12. Colleoni M, Bagnardi V, Rotmensz N, Viale G, Mastropasqua M, Veronesi P, Cardillo A, Torrisi R, Luini A, Goldhirsch A (2010) A nomogram based on the expression of Ki-67, steroid hormone receptors status and number of chemotherapy courses to predict pathological complete remission after preoperative chemotherapy for breast cancer. Eur J Cancer 46:2216–2224. doi:10.1016/j.ejca.2010.04.008
13. Rouzier R, Pusztai L, Delaloge S, Gonzalez-Angulo AM, Andre F, Hess KR, Buzdar AU, Garbay JR, Spielmann M, Mathieu MC, Symmans WF, Wagner P, Atallah D, Valero V, Berry DA, Hortobagyi GN (2005) Nomograms to predict pathologic complete response and metastasis-free survival after preoperative chemotherapy for breast cancer. J Clin Oncol 23:8331–8339. doi:10.1200/JCO.2005.01.2898
14. Lee JK, Coutant C, Kim YC, Qi Y, Theodorescu D, Symmans WF, Baggerly K, Rouzier R, Pusztai L (2010) Prospective comparison of clinical and genomic multivariate predictors of response to neoadjuvant chemotherapy in breast cancer. Clin Cancer Res 16:711–718. doi:10.1158/1078-0432.CCR-09-2247
15. Tabchy A, Valero V, Vidaurre T, Lluch A, Gomez H, Martin M, Qi Y, Barajas-Figueroa LJ, Souchon E, Coutant C, Doimi FD, Ibrahim NK, Gong Y, Hortobagyi GN, Hess KR, Symmans WF, Pusztai L (2010) Evaluation of a 30-gene paclitaxel, fluorouracil, doxorubicin, and cyclophosphamide chemotherapy response predictor in a multicenter randomized trial in breast cancer. Clin Cancer Res 16:5351–5361. doi:10.1158/1078-0432.CCR-10-1265
16. Albain KS, Barlow WE, Shak S, Hortobagyi GN, Livingston RB, Yeh IT, Ravdin P, Bugarini R, Baehner FL, Davidson NE, Sledge GW, Winer EP, Hudis C, Ingle JN, Perez EA, Pritchard KI, Shepherd L, Gralow JR, Yoshizawa C, Allred DC, Osborne CK, Hayes DF (2010) Prognostic and predictive value of the 21-gene recurrence score assay in postmenopausal women with node-positive, oestrogen-receptor-positive breast cancer on chemotherapy: a retrospective analysis of a randomised trial. Lancet Oncol 11:55–65. doi:10.1016/S1470-2045(09)70314-6
17. Liedtke C, Hatzis C, Symmans WF, Desmedt C, Haibe-Kains B, Valero V, Kuerer H, Hortobagyi GN, Piccart-Gebhart M, Sotiriou C, Pusztai L (2009) Genomic grade index is associated with response to chemotherapy in patients with breast cancer. J Clin Oncol 27:3185–3191. doi:10.1200/JCO.2008.18.5934
18. Parker JS, Mullins M, Cheang MC, Leung S, Voduc D, Vickery T, Davies S, Fauron C, He X, Hu Z, Quackenbush JF, Stijleman IJ, Palazzo J, Marron JS, Nobel AB, Mardis E, Nielsen TO, Ellis MJ, Perou CM, Bernard PS (2009) Supervised risk predictor of breast cancer based on intrinsic subtypes. J Clin Oncol 27:1160–1167. doi:10.1200/JCO.2008.18.1370
19. Straver ME, Glas AM, Hannemann J, Wesseling J, van de Vijver MJ, Rutgers EJ, Vrancken Peeters MJ, van Tinteren H, Van't Veer LJ, Rodenhuis S (2010) The 70-gene signature as a response predictor for neoadjuvant chemotherapy in breast cancer. Breast Cancer Res Treat 119:551–558. doi:10.1007/s10549-009-0333-1
20. Freund Y, Mason L (1999) The alternating decision tree learning algorithm. In: Proceedings of the sixteenth international conference on machine learning, pp 124–133
21. Che D, Hockenbury C, Marmelstein R, Rasheed K (2010) Classification of genomic islands using decision trees and their ensemble algorithms. BMC Genomics 11(Suppl 2):S1. doi:10.1186/1471-2164-11-S2-S1
22. von Minckwitz G, Rezai M, Loibl S, Fasching PA, Huober J, Tesch H, Bauerfeind I, Hilfrich J, Eidtmann H, Gerber B, Hanusch C, Kühn T, du Bois A, Blohmer JU, Thomssen C, Dan Costa S, Jackisch C, Kaufmann M, Mehta K, Untch M (2010) Capecitabine in addition to anthracycline- and taxane-based neoadjuvant treatment in patients with primary breast cancer: phase III GeparQuattro study. J Clin Oncol 28:2015–2023. doi:10.1200/JCO.2009.23.8303
23. Jinno H, Sakata M, Hayashida T, Takahashi M, Sato T, Seki H, Kitagawa Y (2011) Primary systemic chemotherapy of breast cancer: indication and predictive factors. Breast Cancer 18:74–79. doi:10.1007/s12282-010-0226-5
24. von Minckwitz G, Blohmer JU, Raab G, Löhr A, Gerber B, Heinrich G, Eidtmann H, Kaufmann M, Hilfrich J, Jackisch C, Zuna I, Costa SD (2005) In vivo chemosensitivity-adapted preoperative chemotherapy in patients with early-stage breast cancer: the GEPARTRIO pilot study. Ann Oncol 16:56–63. doi:10.1093/annonc/mdi001
25. Perou CM (2010) Molecular stratification of triple-negative breast cancers. Oncologist 15(Suppl 5):39–48. doi:10.1634/theoncologist.2010-S5-39
26. Rouzier R, Pusztai L, Garbay JR, Delaloge S, Hunt KK, Hortobagyi GN, Berry D, Kuerer HM (2006) Development and validation of nomograms for predicting residual tumor size and the probability of successful conservative surgery with neoadjuvant chemotherapy for breast cancer. Cancer 107:1459–1466. doi:10.1002/cncr.22177
27. Fan C, Prat A, Parker JS, Liu Y, Carey LA, Troester MA, Perou CM (2011) Building prognostic models for breast cancer patients using clinical variables and hundreds of gene expression signatures. BMC Med Genomics 4:3. doi:10.1186/1755-8794-4-3